A Survey on Food Computing

WEIQING MIN, SHUQIANG JIANG, and LINHU LIU, Key Lab of Intelligent Information Processing, Institute of Computing Technology, CAS, China
YONG RUI, Lenovo Group, China
RAMESH JAIN, Department of Computer Science, University of California, USA

Food is essential for human life and it is fundamental to the human experience. Food-related study may support multifarious applications and services, such as guiding human behavior, improving human health, and understanding the culinary culture. With the rapid development of social networks, mobile networks, and Internet of Things (IoT), people commonly upload, share, and record food images, recipes, cooking videos, and food diaries, leading to large-scale food data. Large-scale food data offers rich knowledge about food and can help tackle many central issues of human society. Therefore, it is time to group several disparate issues related to food computing. Food computing acquires and analyzes heterogenous food data from different sources for perception, recognition, retrieval, recommendation, and monitoring of food. In food computing, computational approaches are applied to address food-related issues in medicine, biology, gastronomy, and agronomy. Both large-scale food data and recent breakthroughs in computer science are transforming the way we analyze food data. Therefore, a series of works has been conducted in the food area, targeting different food-oriented tasks and applications. However, there are very few systematic reviews that shape this area well and provide a comprehensive and in-depth summary of current efforts or detail open problems in this area. In this article, we formalize food computing and present such a comprehensive overview of various emerging concepts, methods, and tasks. We summarize key challenges and future directions ahead for food computing. This is the first comprehensive survey that targets the study of computing technology for the food area and also offers a collection of research studies and technologies to benefit researchers and practitioners working in different food-related fields.

CCS Concepts: • **General and reference → Surveys and overviews**; • **Information systems → Multimedia information systems**; *Information retrieval*; • **Applied computing → Health care information systems**;

Additional Key Words and Phrases: Food computing, food recognition, health, food perception, food retrieval, recipe analysis, recipe recommendation, monitoring, survey

92

This work was supported in part by the National Natural Science Foundation of China (No. 61972378, 61602437, and 61532018), in part by the Lenovo Outstanding Young Scientists Program, in part by the National Program for Special Support of Eminent Professionals and National Program for Support of Top-notch Young Professionals, and in part by the China Postdoctoral Science Foundation (2017T100110).
Authors' addresses: W. Min, S. Jiang (corresponding author), and L. Liu, Key Lab of Intelligent Information Processing, Institute of Computing Technology, CAS, Beijing, China; emails: {minweiqing, sqjiang}@ict.ac.cn, linhu.liu@vipl.ict.ac.cn; Y. Rui, Lenovo Group, No. 6, Shangdi West Road, Beijing, China; email: yongrui@lenovo.com; R. Jain, Department of Computer Science, University of California, Irvine, CA; email: jain@ics.uci.edu.

0360-0300/2019/09-ART92 $15.00
https://doi.org/10.1145/3329168

ACM Reference format:
Weiqing Min, Shuqiang Jiang, Linhu Liu, Yong Rui, and Ramesh Jain. 2019. A Survey on Food Computing. *ACM Comput. Surv.* 52, 5, Article 92 (September 2019), 36 pages.
https://doi.org/10.1145/3329168

1 INTRODUCTION

Food has a profound impact on human life, health, and wellbeing (Achananuparp et al. 2018; Nordstrom et al. 2013). An increasing amount of people are becoming overweight or obese. According to WHO, there are more than 1.9B adults aged 18 or over who are overweight, whereas there are more than 650M people who are obese. The worldwide prevalence of obesity in 2016 is nearly three times that of 1975.[1] Being overweight and obesity have been found to be major risk factors for various chronic diseases, such as diabetes and cardiovascular diseases.[2] For example, it is estimated that 415M people suffer from diabetes worldwide in 2015.[3] One important reason is that many generally maintain an excessively unhealthy lifestyle and bad dietary habits (Ng et al. 2014), such as the increased intake of food with high energy and high fat. In addition, food is much more than a tool for survival. It plays an important role in defining our identity, social status, religious significance, and culture (Harris 1985; Khanna 2009); as Jean Anthelme Brillat-Savarin said, "Tell me what you eat, and I will tell you who you are." Furthermore, how we cook and how we eat food are also factors profoundly touched by our individual cultural inheritance. For these reasons, food-related study (Ahn et al. 2011; Bucher et al. 2013; Canetti et al. 2002; Chung et al. 2017; Sajadmanesh et al. 2017) has always been a hot topic and received extensive attention from various fields.

In earlier years, food-related study was conducted from different aspects, such as food choice (Nestle et al. 1998), food perception (Sorensen et al. 2003), food consumption (Pauly 1986), food safety (Chen and Tao 2001), and food culture (Harris 1985). However, these studies were conducted using traditional approaches before the web revolutionized research in many areas. In addition, most methods use small-scale data, such as questionnaires, cookbooks, and recipes. Nowadays, the fast development of various networks, such as social networks, mobile networks, and IoT, allows users to easily share food images, recipes, cooking videos, or record food diaries via these networks, leading to large-scale food datasets. These food data imply rich knowledge and thus can provide great opportunities for food-related study, such as discovering principles of food perception (Mouritsen et al. 2017), analyzing culinary habits (Sajadmanesh et al. 2017), and monitoring the diet (Chung et al. 2017). In addition, various new data analysis methods in network analysis, computer vision, machine learning, and data mining are proposed. Recent breakthroughs in Artificial Intelligence (AI), especially deep learning (LeCun et al. 2015), have further fueled the interest in large-scale food-oriented study (Chen et al. 2017c; Hassannejad et al. 2016; Kawano and Yanai 2014b; Pandey et al. 2017) for its superiority in learning representations from various types of signals.

Taking these factors into consideration, we come up with a vision of food computing that aims to apply heterogeneous food data collected from different data sources to various applications in different fields. To our knowledge, Harper and Siller (2015) first proposed the term "food computing" in the agricultural field. However, they did not give clear definition. In a broad sense, we think that food computing focuses on food-related study via computer science, and it is an interdisciplinary field. Consequently, there are many open questions to answer. For example, what

[1]http://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight.
[2]http://www.who.int/mediacentre/factsheets/fs311/en/index.html.
[3]http://www.diabetesatlas.org/.

are the core research problems of food computing? What are the key methodologies for food computing? What are representative applications in this domain? What are challenges and potential directions for this research field?

To answer these questions, we formally define food computing in this article and introduce its general framework, tasks, and applications. Some food-related surveys have been done. For example, Knez and Šajn (2015) gave a survey on mobile food recognition, and nine recognition systems are introduced based on their system architecture. Trattner and Elsweiler (2017a) provided a summary of food recommender systems. BVR and J (2017) presented a variety of methodologies and resources on automatic food monitoring and diet management system. However, to the best of our knowledge, there are very few systematic reviews that shape this area well and provide a comprehensive and in-depth summary of current efforts, challenges, or future directions in the area. This survey seeks to provide such a comprehensive summary of current research on food computing to identify open problems and point out future directions. It aims to build the connection between computer science and food-related fields, serving as a good reference for developing food computing techniques and applications for various food-related fields. To this end, about 300 studies are shortlisted and classified in this survey.

This survey is organized as follows: Section 2 first presents the concept and framework of food computing. Section 3 introduces food data acquisition and analysis, where different types of food datasets are summarized and compared. We present its representative applications in Section 4. Main tasks in food computing are reviewed in Section 5. Sections 6 and 7 discuss its challenges and prominent open research issues, respectively. We conclude the article in Section 8.

2 FOOD COMPUTING

Food computing mainly utilizes the methods from computer science for food-related study. It involves the acquisition and analysis of food data with different modalities (e.g., food images, food logs, recipe, taste, and smell) from different data sources (e.g., the social network, recipe-sharing websites, and cameras). Such analysis resorts to computer vision, machine learning, data mining, and other advanced technologies to connect food and humans for supporting human-centric services, such as human behavior and health. It is a typically multidisciplinary field, where computer science meets conventional food-related fields, such as food science, medicine, biology, agronomy, sociology, and gastronomy. Therefore, besides computer science, food computing also borrows theories and methods from other disciplines, such as neuroscience, cognitive science, and chemistry. As shown in Figure 1, food computing mainly consists of five basic tasks, from perception, recognition, retrieval, and recommendation to prediction and monitoring. It further enables applications for various fields.

Food computing applies computational approaches for acquiring and analyzing heterogenous food data from disparate sources for perception, recognition, retrieval, recommendation, and monitoring of food to address food-related issues in health, biology, gastronomy, and agronomy.

Figure 1 shows its general framework. One important goal of food computing is to provide various human-centric services. Therefore, the first step is to collect human-produced food data. We can acquire food data of different types from various data sources. In addition, there are also other specific food datasets available, such as the odor threshold database and the volatile compounds in food database. Based on these food data, we utilize different technologies, such as machine learning and computing vision for food data analysis. After that, we can conduct five main food computing tasks. The flavor and sensory perception of food can govern our choice of food and affect how much we eat or drink. Food perception is multi-modal, including visual information, tastes, smells, and tactile sensations. Recognition is one basic task and it is mainly to predict food items such as the category or ingredients from food images. Food-oriented retrieval involves

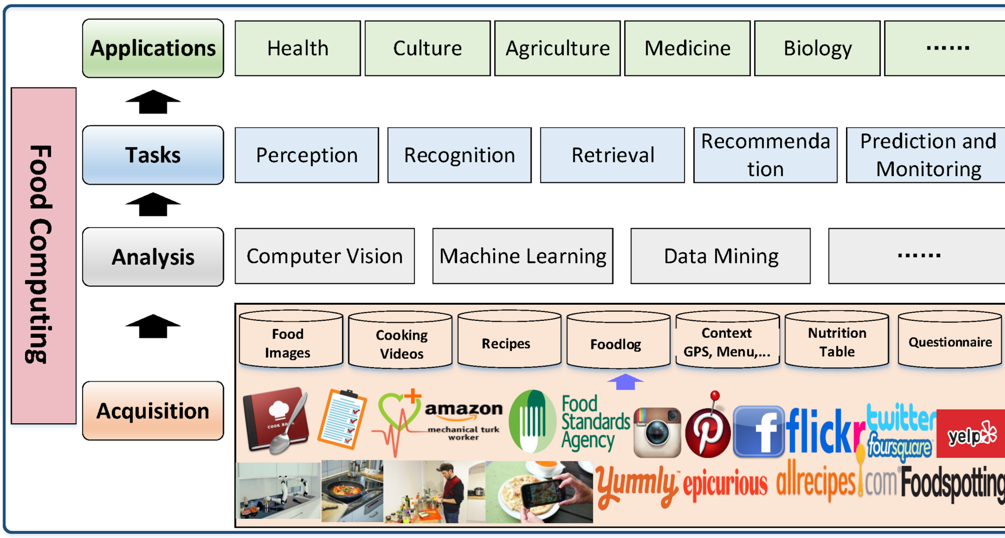

Fig. 1. An overview of food computing.

single-modality based retrieval (such as visual food retrieval and recipe retrieval) and cross-modal retrieval, which receives more attention for its applications such as retrieving recipes from food images. Food-oriented recommendation can not only recommend the food people might want to eat, but also provide them with a healthier diet. Food recommendation involves more complex and multi-faceted information. Therefore, it is different from other types of recommendations. Prediction and monitoring are mainly conducted based on social media, such as monitoring public health.

Furthermore, different tasks are not independent but closely intertwined and mutually dependent. For example, the recognized results can further support retrieval, recommendation, and even food perception. When the categories of food images are huge, retrieval-based methods can also be used for food recognition. Prediction from social media can be helpful for the recommendation task. For example, users' food preferences predicted from social media is an important step towards personalized food recommendation.

3 FOOD DATA ACQUISITION AND ANALYSIS

In this section, we introduce frequently used data in food computing and briefly give the summary and comparison on existing food datasets.

Benefiting from the development of the internet and various smart devices, a number of research works focus on studying food perception, pattern mining, and human behavior via various data-driven methods (Mouritsen et al. 2017). For example, to analyze a user's eating habits for his/her dietary assessment, we should acquire his/her food log data for further analysis. Through the analysis of these food data, we can discover some general principles that may underlie food perception and diverse culinary practice. Therefore, the first step of food computing involves the acquisition and collection of food data. Particularly, we summarize existing data sources into three main types: (1) websites, (2) social media, and (3) cameras.

In the early years, researchers mainly obtained food data from official organizations to conduct food-related study. For example, Sherman and Billing (1999) analyzed 93 traditional cookbooks from 36 countries to find the reason that humans use spices. To calculate the food calorie, they

Fig. 2. (a) Some recipes from Yummly (b) Food images from two datasets: ETHZ Food-101 (Bossard et al. 2014) and UEC Food256 (Kawano and Yanai 2014a).

should search its energy in the nutrition table provided by official organizations, e.g., United States Department of Agriculture (USDA)[4] and BLS.[5] These data acquisition methods are generally time-consuming, laborious, and hard to achieve the large-scale.

The proliferation of recipe-sharing websites has resulted in huge online food data collections. These websites, such as Yummly, Meishijie, foodspotting. and Allrecipes, have emerged over the past several years. Besides basic information, such as lists of ingredients, these recipes are associated with rich modality and attribute information. Figure 2(a) shows some examples from Yummly. Each recipe includes a list of ingredients, food image, cuisine category, course, flavor, and macronutrient composition. Such recipe data with rich types can be exploited to answer various food-related questions, such as pattern analysis on ingredient combination from different regions (Ahn et al. 2011; Min et al. 2018) and food recognition (Bossard et al. 2014). As one representative work, Sajadmanesh et al. (2017) built a large-scale recipe dataset from Yummly with 157,013 recipes from over 200 types of cuisines for culinary habit analysis. In addition, there is rich social information provided by some recipe websites, e.g., ratings and comments, that can be helpful for tasks such as recipe recommendation (Teng et al. 2012) and recipe rating prediction (Yu et al. 2013).

Besides recipe-sharing websites, social media, such as Twitter, Facebook, Foursquare, Flickr, Instagram, and YouTube also provide large-scale food data. For example, Culotta (2014) examined whether linguistic patterns in Twitter correlate with health-related statistics. Abbar et al. (2015) combined demographic information and food names from Twitter to model the correlation between calorie value and diabetes. In addition to textual data, recent studies (Mejova et al. 2016;

[4] https://ndb.nal.usda.gov/ndb/.
[5] https://www.blsdb.de.

Ofli et al. 2017) have used large-scale food images from social media for the study of food perception and eating behaviors.

With the popularity of cameras embedded in smartphones and various wearable devices (Vu et al. 2017), collecting food data directly from cameras is also a common way. For example, researchers have begun capturing food images in restaurants or canteens for visual food understanding (Ciocca et al. 2016; Damen et al. 2018). Besides food images, Damen et al. (2018) used the head-mounted GoPro camera to collect cooking videos.

In summary, the types of food-related data from different sources are divided into the following types:

- Recipes: Recipes contain a set of ingredients and sequential cooking instructions. In earlier research, recipes were collected from cookbooks and manually typed into computers. Currently, recipes can be collected from recipe websites, such as epicurious and Allrecipes. As a result, their numbers have grown exponentially. Such types of data can be embedded in the latent space for recipe analysis to further support various applications (Kim and Chung 2016).
- Dish images: Dish images are the most common multimedia data with rich visual information and semantic content. We can extract meaningful concepts and information to support various applications. Most tasks conduct the visual analysis for food images with the single item. There are also some food image datasets such as UEC Food256 (Kawano and Yanai 2014a) and UNIMIB2016 (Ciocca et al. 2016) with multiple food-items. Figure 2(b) shows some examples.
- Cooking videos: Nowadays, there are plenty of cooking videos that can guide people how to cook. They contain cooking activities and cooking procedure information. Researchers can use such data for cooking activity recognition and other tasks (Damen et al. 2018).
- Food attributes: Food contains rich attributes, such as flavors, cuisine, taste, smell, cooking, and cutting attributes. We can adopt rich food attributes to improve food recognition and other tasks (Chen et al. 2017a; Min et al. 2017a).
- Foodlog: Foodlog records food images, text, and other calorie information. With the rapid growth of mobile technologies and applications, we can use the Foodlog app to keep a healthy diet. Some works, such as Kitamura et al. (2008), introduced a food-logging system for food balance estimation.
- Restaurant-relevant food information: Nowadays, more works use restaurant-specific information, such as the menu and GPS information for restaurant-specific food recognition (Herranz et al. 2017) and further food logging (Beijbom et al. 2015).
- Healthiness: More people pay attention to health because of an improved living standard. Healthiness contains rich information, such as the calorie and nutrition. An excessively unhealthy lifestyle and bad dietary habits can trigger obesity and other diseases. Researchers can use the healthiness of food for automatic food calorie estimation from food images to keep a healthy diet (Okamoto and Yanai 2016).
- Other food data: Other food data include the data from cookbooks, questionnaires, odor threshold database,[6] and food product codes. The data by questionnaire (Thompson et al. 2008) includes diverse forms, such as Food Frequency Questionnaires (FFQ) and Food Cravings Questionnaire (FCQ).

After obtaining the initial food collection, especially from web and social media, the next step is data annotation. One simple way is to directly utilize tags from websites or social media as the

[6] http://www.thresholdcompilation.com/.

Table 1. Food-related Datasets

Reference	Dataset Name	Data Type	Num.	Sources	Tasks
(Chen et al. 2009)	PFID	Images with categories	4,545 (101)	Cameras	Recognition
(Joutou and Yanai 2010)	Food50	Images with categories	5K (50)	Web	Recognition
(Hoashi et al. 2010)	Food85	Images with categories	8,500 (85)	Web	Recognition
(Chen et al. 2012)	-	Images with categories	5K (50)	Web + Cameras	Quantity Estimation
(Matsuda and Yanai 2012)	UEC Food100[1]	Images with categories	14,361 (100)	Web	Recognition
(Anthimopoulos et al. 2014)	Diabetes	Images with categories	4,868 (11)	Web	Recognition
(Kawano and Yanai 2014a)	UEC Food256[2]	Images with categories	25,088 (256)	Web	Recognition
(Bossard et al. 2014)	ETHZ Food-101[3]	Images with categories	101K (101)	Web (foodspotting)	Recognition
(Wang et al. 2015)	UPMC Food-101[4]	Images and text with categories	90,840 (101)	Web (Google search)	Recognition
(Farinella et al. 2014a)	UNICT-FD889[5]	Images with categories	3,583 (889)	Cameras (Smartphone)	Retrieval
(Pouladzadeh et al. 2015)	FooDD[6]	Images with categories	3K (23)	Camera	Detection
(Meyers et al. 2015)	Food201-Segmented	Images with categories	12,625 (201)	Web (e.g., Flickr, Instagram)	Segmentation
(Bettadapura et al. 2015)	-	Images with categories and location	3,750 (75)	Cameras	Recognition
(Xu et al. 2015)	Dishes[7]	Images with categories and location	117,504 (3,832)	Web (Dianping)	Recognition
(Beijbom et al. 2015)	Menu-Match[8]	Images with categories	646(41)	Cameras (Smartphone, Instamatic)	Food Logging
(Ciocca et al. 2015)	UNIMIB2015[9]	Images with categories	2K (15)	Cameras (Smartphone)	Recognition
(Ciocca et al. 2016)	UNIMIB2016[9]	Images with categories	1,027 (73)	Cameras (Smartphone)	Recognition
(Zhou and Lin 2016)	Food-975	Images with categories	37,785 (975)	Camera + Web(yelp)	Recognition
(Merler et al. 2016)	Food500	Images with categories	148,408 (508)	Web(e.g., Bing) + Social media (Instagram)	Recognition
(Rich et al. 2016)	Instagram800K[10]	Images with tags	808,964 (43)	Social media (Instagram)	Recognition
(Singla et al. 2016)	Food11	Images with categories	5K (50)	Social media (e.g., Flickr)	Recognition
(Farinella et al. 2016)	UNICT-FD1200[11]	Images with categories	4,754 (1,200)	Cameras (Smartphone)	Recognition and Retrieval
(Ofli et al. 2017)	-	Images with tags	1.9M	Social media (Instagram)	Food Perception
(Liang and Li 2017)	ECUSTFD[12]	Images with rich annotation	2,978 (19)	Camera (Smartphone)	Calorie Estimation
(Ciocca et al. 2017)	Food524DB[13]	Images with categories	247,636 (524)	Web + Camera	Recognition
(Chen et al. 2017e)	ChineseFoodNet[14]	Images with categories	192K (208)	Web + Camera	-
(Thanh and Gatica-Perez 2017)	Instagram 1.7M	Images with comments	1.7M	Social media (Instagram)	Consumption Patterns Analysis
(Harashima et al. 2017)	Cookpad[15]	Images and recipes	4,748,044	Web(Cookpad)	-

[1]http://foodcam.mobi/dataset100.html/. [2]http://foodcam.mobi/dataset256.html/. [3]http://www.vision.ee.ethz.ch/datasets_extra/food-101/. [4]http://visiir.lip6.fr/. [5]http://iplab.dmi.unict.it/UNICT-FD889/. [6]http://www.site.uottawa.ca/~shervin/food/. [7]http://isia.ict.ac.cn/dataset/Geolocation-food/. [8]http://neelj.com/projects/menumatch/. [9]http://www.ivl.disco.unimib.it/activities/food-recognition/. [10]http://www.eecs.qmul.ac.uk/tmh/downloads.html. [11]http://www.iplab.dmi.unict.it/UNICT-FD1200/. [12]https://github.com/Liang-yc/ECUSTFD-resized-. [13]http://www.ivl.disco.unimib.it/activities/food524db/. [14]https://sites.google.com/view/chinesefoodnet/. [15]https://www.nii.ac.jp/dsc/idr/cookpad/cookpad.html.

annotation. However, such annotations are probably noisy. One probable way is manual annotation by ourselves or nutrition experts (Martin et al. 2012). However, such a method is limited to small-scale data. To annotate large-scale data, crowd-sourcing is generally used, e.g., Amazon Mechanical Turk (AMT) (Kawano and Yanai 2014a).

Existing Benchmark Food Datasets. Many benchmark and popular food datasets have been constructed and released. Table 1 list main food-related databases in more details, where the number in () denotes the number of categories for the column Num, and particular websites or cameras for the column of Sources. We also give the links for datasets if available. From Table 1, we can see that: (1) The benchmark datasets for food recognition are released frequently. Earlier, researchers focus on the food dataset with few cuisines and small-scale. For example, UEC Food100 (Matsuda

Table 1. Continued

Reference	Dataset Name	Data Type	Num.	Sources	Tasks
(Rohrbach et al. 2012)	MPII Cooking 2[16]	Cooking videos	273	Cameras	Cooking Activity Recognition
(Stein and Mckenna 2013)	50 Salads[17]	Cooking videos	50	Cameras	Cooking Activity Recognition
(Kuehne et al. 2014)	Breakfast[18]	Cooking videos	433	Cameras	Cooking Activity Recognition
(Damen et al. 2018)	EPIC-KITCHENS[19]	Cooking videos	432	Cameras(GoPro)	Cooking Activity Recognition
(Kinouchi et al. 2008)	-	Recipes	7,702	-	Culinary Evolution
(Ahn et al. 2011)	Recipes56K[20]	Recipes	56,498	Web	Ingredient Pattern Discovery
(Teng et al. 2012)	-	Recipes	46,337	Web (allrecipes)	Recipe Recommendation
(Kim and Chung 2016)	-	Recipes	5,917	Web (Recipesource)	Recipe Analysis
(Chen and Ngo 2016)	Vireo Food-172[21]	Recipes with images and ingredients	110,241(172)	Web	Recipe Retrieval
(Sajadmanesh et al. 2017)	Recipes157K	Recipes with metadata	157K	Web (Yummly)	Cross-region Food Analysis
(Chen et al. 2017b)	Go cooking	Recipes&Images	61,139	Web (xiachufang)	Cross-modal Recipe Retrieval
(Salvador et al. 2017)	Recipe1M[22]	Recipes&Images	1M	Web	Cross-modal Recipe Retrieval
(Min et al. 2017a)	Yummly-28K[23]	Recipes&Images	28K	Web (Yummly)	Cross-modal Retrieval
(Min et al. 2018)	Yummly-66K[24]	Recipes&Images	66K	Web (Yummly)	Cross-region Food Analysis
(Markus et al. 2018)	Recipes242K[25]	Recipes	242,113	Web (Allrecipes)	Recipe Healthiness Estimation
(Semih et al. 2018)	RecipeQA[26]	Recipes	20K(22)	Web (Instructables)	Recipe Question Answering

[16]https://www.mpi-inf.mpg.de/departments/computer-vision-and-multimodal-computing/research/human-activity-recognition/mpii-cooking-2-dataset/. [17]http://cvip.computing.dundee.ac.uk/datasets/foodpreparation/50salads/. [18]http://serre-lab.clps.brown.edu/resource/breakfast-actions-dataset/#Downloads. [19]https://epic-kitchens.github.io/2018. [20]http://www.yongyeol.com/2011/12/15/paper-flavor-network.html. [21]http://vireo.cs.cityu.edu.hk/VireoFood172/. [22]http://im2recipe.csail.mit.edu/. [23]http://isia.ict.ac.cn/dataset/. [24]http://isia.ict.ac.cn/dataset/Yummly-66K.html. [25]https://github.com/rokickim/nutrition-prediction-dataset/blob/master/. [26]https://hucvl.github.io/recipeqa.

and Yanai 2012) consists of 14,361 Japanese food images. Benefiting from the fast development of social media and mobile devices, we can easily obtain more food images. For example, Rich et al. (2016) released a dataset with 808,964 images from Instagram. In addition, ETHZ Food-101 (Bossard et al. 2014) has been a benchmark food dataset for the food recognition task. (2) There are some restaurant-oriented datasets, such as Dishes (Xu et al. 2015) and Menu-Match (Beijbom et al. 2015). Such datasets generally contain the location information, such as GPS or restaurant information. (3) Compared with food images, recipes contain richer attributes and metadata information. To the best of our knowledge, Recipe1M (Salvador et al. 2017) is the largest released recipe dataset, with 1M cooking recipes and 800K images. Recently, Semih et al. (2018) released a recipe dataset RecipeQA, which includes an additional 36K questions to support question answering compared with other recipe datasets. Some datasets with cooking videos have also been released for human-activity recognition and prediction, e.g., recently released EPIC-KITCHENS (Damen et al. 2018).

Summary and Discussion. In this section, we summarized existing food-related data sources into three main types, namely websites (e.g., Yummly, Meishijie, foodspotting, and Allrecipes), social media (e.g., Twitter, Facebook, Foursquare, Flickr, Instagram, and YouTube), and cameras (e.g., smartphone and point-and-shoot camera). We also listed different types of food data, such as recipes, dish images, and food attributes, and, finally, compared existing food datasets. After initial food data collection, different annotation methods are introduced, such as tags from social media and websites, manual annotation, or crowd-sourcing.

These increasing amounts of food-related data present researchers with more opportunities for food analysis. Such analysis can be conducted not only on these datasets individually, but also multiple datasets jointly. For example, we can analyze the correlation between chemical data and recipes (Ahn et al. 2011) or social media images and obesity (Mejova et al. 2016). These connections with different kinds of food data can provide us with a new perspective on the study of food from different angles, such as the culinary habits and human behavior.

4 APPLICATIONS IN FOOD COMPUTING

Before introducing core tasks in food computing, we first list a number of applications and summarize them from the following four main aspects: health, agriculture, culture, and food science.

4.1 Health

What kind of food or how much we eat is closely related to our health. For example, if we eat too much, we can risk developing multiple types of diseases, such as diabetes and heart disease. Therefore, food-relevant study will benefit various health-oriented applications. Particularly, we introduce four representative food-oriented health applications, including (1) food perception for health, (2) food recognition for diet management, (3) health-aware food recommendation, and (4) food-health analysis from social media.

Food Perception for Health. One important aspect determining our food choices and how much we eat/drink is how we perceive food from its certain characteristics, such as whether it is sweet or tasty. An increasing number of researchers studied how we perceive food, both before and during its consumption, and have proved the influence of sensory properties of food on eating behavior (Sorensen et al. 2003). In addition, multimodal sensory cues can affect the food identification and the guidance of food choice (McCrickerd and Forde 2016).

Dietary Management for Health. Dietary assessment or food diary (Achananuparp et al. 2018; Cordeiro et al. 2015a, 2015b) provides valuable insights for disease prevention. With the advancement of smart devices and computer vision technologies, more approaches utilize vision methods to process food photos captured by smartphone for diet management. To our knowledge, the first attempt for food intake analysis from the photo is to measure the food intake in the cafeteria settings, developed by Williamson et al. (2003). This method is semi-automatic and involves the participation of registered dietitians. To make the system fully automatic, Zhu et al. (2010) proposed a dietary assessment system, where images obtained before and after food is eaten are used to estimate the category and amount of consumed food. Similar methods including single-view reconstruction and multi-view reconstruction for food volume estimation (Dehais et al. 2017; Pouladzadeh et al. 2014) are proposed. Recently, a lot of works focus on calorie estimation from one image (Fang et al. 2018; Meyers et al. 2015). As representative work, Meyers et al. (2015) proposed an Im2Calories system, which first localized the meal region from one food photo, and then labeled these segmented regions and estimated their volume. In addition, more works conducted food calorie estimation on mobile devices (BVR and J 2017; Pouladzadeh et al. 2016b) and other wearable devices, such as Kinect and glasses with load cells (Vu et al. 2017). Recently, researchers designed new sensors to track the diets and count the calories (Strickland 2018).

Health-aware Food Recommendation. Many people are facing the problem of making healthier food decisions to reduce the risk of chronic diseases such as obesity and diabetes, which are very relevant to what we eat. Therefore, food recommendation not only caters to users' food preferences but should be also able to take users' health into account, leading to heath-aware food recommendation. The core problem for health-aware food recommendation is to build the model to balance these two components, and thus is helpful for healthy diet. Recently, many works focus on this topic. For example, Yang et al. (2017) learned users' preferences from a large food image dataset and projected these preferences for general food items into the domain that meets each individual user's health goals. Considering huge potentials in human health, we will see the surge in the health-aware food recommendation field.

Food-Health Analysis from Social Media. We're in an era of social media. As food is indispensable to our life, a great deal of online content is relevant to food. Therefore, a great amount of food information about our culinary habits and behavior from social media can be explored for food-health analysis. Recent studies have shown that we can use social media to get aggregated statistics about the health of people, such as the health insurance coverage and obesity for public health monitoring (Culotta 2014; Mejova et al. 2016).

4.2 Culture

Food is fundamental to the culture, with food practices reflecting our nationalities and other aspects (Bell 1997; Giampiccoli and Kalis 2012; Harris 1985; Khanna 2009). An understanding of food culture is indispensable in human communication. This is true not only for professionals in fields such as public health and commercial food services, but is clearly recognized in the global marketplace. Food has also come to be recognized as part of the local culture that tourists consume, as an element of regional tourism promotion and a potential component of local agricultural and economic development (Hall and Hall 2003). In addition, exploring the food culture can help develop personalized food recommendation considering the aspect of food culture from different urban areas.

For these reasons, the study of culinary cultures began to receive more attention (Ahn et al. 2011; Kim and Chung 2016; Sajadmanesh et al. 2017; Zhu et al. 2013). Ahn et al. (2011) identified significant ingredient patterns that indicate the way humans choose paired ingredients in their food. These patterns vary from geographic region to geographic region. For example, the ingredients with shared flavor compounds tend to be combined for North American dishes. Sajadmanesh et al. (2017) further analyzed and compared worldwide cuisines and culinary habits using a larger recipe dataset. However, these works only mined recipe text for analysis and ignored rich visual information. Min et al. (2018) recently combined food images with recipes from Yummly for multimodal cuisine summarization to further analyze the culinary cultures. The visual information enables the analysis and comparison of culinary cultures easily and more comprehensively. Besides recipes, social media–based food culture analysis has been conducted, such as dietary choice study (Abbar et al. 2015; Ofli et al. 2017). The prosperity of social media provides opportunities to obtain detailed and complete records of individual food consumption, which will continue revolutionizing the way we understand the culinary culture.

4.3 Agriculture

Food computing can also be used in the agriculture or food products. Food image analysis has great potential for automated agricultural and food safety tasks (Senthilnath et al. 2016; Xiang et al. 2014). For example, Jimenez et al. (1999) proposed a recognition system to locate the fruit. Recently, artificial vision systems (Chen et al. 2017c; Hernandez-Hernandez et al. 2017; Lu et al. 2017) have become powerful tools for automatic recognition of fruits and vegetables, because of its powerful capacity of feature representation. For example, Hernandez-Hernandez et al. (2017) presented an image capture, cropping, and process for fruit recognition. Chen et al. (2017c) introduced a deep learning method to extract visual features for counting fruits. In addition, there are some works for natural food product classification, such as tomato ripeness classification (Pabico et al. 2015) and rice variety classification (Chatnuntawech et al. 2018). Chatnuntawech et al. (2018) developed a non-destructive system that first used a hyperspectral imaging system to acquire complementary spatial and spectral information of rice seeds and then used Convolutional Neural Networks (CNNs) (Krizhevsky et al. 2012) to extract features from spatio-spectral data to determine the rice varieties.

It is worth noting that agriculture-oriented food recognition is more similar to visual object recognition, such as fruit recognition. However, it is quite different from dish or ingredient recognition. In contrast to object-like recognition, food typically does not exhibit any distinctive semantic parts. As a result, we should design new recognition methods or paradigms for dish or ingredient recognition.

4.4 Food Science

According to Wikipedia, food science is defined as the application of basic sciences and engineering to study the physical, chemical, and biochemical nature of foods and principles of food processing.[7] Food computing provides new methods and technologies for these sub-areas. For example, sensory analysis is used to study how human senses perceive food. Food perception uses the Magnetic Resonance Imaging (MRI) to measure brain activity–based perception, and thus is often conducted in the lab (Killgore and Yurgelun-Todd 2005). In contrast, Ofli et al. (2017) considered this problem as food image recognition from Instagram and showed the perception gap between how a machine labels an image and how a human does. In addition, food perception should be multi-modal and it includes visual and auditory cues, tastes, smells, and tactile sensations. Therefore, multi-modal integration is needed. Existing studies (Verhagen and Engelen 2006) focused on this topic from the neuroscience. However, we can resort to deep learning–based multimodal learning methods (Srivastava and Salakhutdinov 2012) in computer science to better tackle this problem. Another example is the quality control. Some works (Pabico et al. 2015) used the neural network to automate the classification of tomato ripeness and acceptability of eggs.

5 TASKS IN FOOD COMPUTING

In this section, we introduce each of five main tasks in turn according to Figure 1.

5.1 Perception

As mentioned before, food perception plays an important part in our health. In addition, such study will have great potential for food and beverage industries; for example, a better understanding of the process used by people to assess the acceptability and flavor of new food products.

Traditional studies on food perception are conducted at the level of brain activity typically in labs. Some works conducted the analysis on the relations between the weight from subjects and food-related stimuli (Killgore et al. 2003; Nenad et al. 2016; Rosenbaum et al. 2008a; Sorensen et al. 2003). For example, Nenad et al. (2016) found that both lean and overweight subjects showed similar patterns of neural responses to some attributes of food, such as smell and taste. There are also some works that are more directly related to visual perception of food. For example, Spence et al. (2010) studied the influence of food color on perceiving the taste and flavor. Ofli et al. (2017) used the image recognition method to study the relation between how food is perceived and what it is actually, namely the food perception gap.

However, our experience of food is multimodal—we not only see food objects, but also hear sounds when chewing, feel its texture, smell its odors, and taste its flavors. Therefore, food perception actually involves multi-modalities. When we are chewing food, we can perceive the taste, flavor, or texture that will facilitate our appreciation of food. The senses of taste and smell play

[7] https://en.wikipedia.org/wiki/Food_science.

a great role in choosing food. Visual information of a food product is essential in the choice and acceptance of this product, while auditory information obtained during the chewing of food products will help us judge whether a product is fresh or not. Food perception does not just depend on one sense, but should be the result from multisensory integration on various types of signals. For example, McCrickerd and Forde (2016) studied the role of multimodal cues including both visual and odor ones in recognizing and selecting food. Particularly, they described the effect of the size of a plate or the amount of food served on the food intake. Verhagen and Engelen (2006) reviewed existing works on multimodal food perception and its neurocognitive bases.

Summary and Discussion. Food perception has received rapid growth of research interest especially in the neuroscience, cognition, and health-related fields. The methodology is being in transition, from neuroscience-based methods in the lab to computational ones. However, advanced computer vision and machine learning methods in computer science have not been fully exploited for food perception. For example, one important problem of multimodal food perception is that of how multimodal features of food are integrated effectively. A feasible method is to employ existing deep networks, such as Srivastava and Salakhutdinov (2012) for effective fusion on heterogeneous signals. Note that recently, some works, such as Ofli et al. (2017), are beginning utilizing big data from websites and social media and computer vision from AI for the study of food perception. The fast development of AI and the increasing availability of food data is likely to result in the establishment of new research disciplines, such as "computational food perception."

5.2 Recognition

The widespread use of smartphones and advances in computer vision enabled novel food recognition systems for dietary assessment, which is a key factor to prevent and treat these diseases. Once we recognize the category or ingredients of the meal, we can further conduct various health-related analysis, e.g., calorie intake estimation, nutrition analysis, and eating habits analysis. In addition, recognizing food directly from images is also highly desirable for other food-related applications. Take self-service restaurants as an example: food recognition can not only monitor the food consumption, but also automatically bill the grabbed meal by the customer. Finally, for people who would like to get a better understanding of food that they are not familiar with or they haven't even seen before, they can simply take a picture and get to know more details about it.

For these reasons, we have seen an explosion in food recognition algorithms in recent years, which are generally divided into the following two types: (1) single-label food recognition, which targets for food images with only one food-item and (2) multi-label food recognition and detection for food images with multiple food-items. In addition, because of wide use in mobile devices and other sensing devices, we also summarize (3) sensor-based food recognition and monitoring. After food recognition, the following step is generally (4) food portion estimation especially in calorie estimation and other dietary management. We finally introduce (5) personalized food recognition for its applications in personal food logging and recommendation.

5.2.1 Single-label Food Recognition. Most research works on food recognition only considered food images with one food item. Relevant works include both hand-crafted and deep representations for multi-class food recognition.

There are two ways using hand-crafted features, single type of features, or the combination of different types. SIFT features (Lowe 2004) are widely used as visual features for food classification (Anthimopoulos et al. 2014; Wu and Yang 2009; Yang et al. 2010). For example, Yang et al. (2010) first employed the semantic texton forest to classify all image pixels into several categories and then obtained the pairwise feature distribution as visual features. In contrast, most methods (Joutou and Yanai 2010; Martinel et al. 2015; Nguyen et al. 2014) combine different types of hand-crafted features to enhance the performance of food recognition. For example, Martinel et al. (2015) used various types of features such as Garbor, LBP and GIST, and then exploited a subset to obtain the optimal ranking performance.

Recently, CNN has been widely used for feature extraction in food recognition and achieves great performance improvement than hand-crafted features. Different types of networks are used in the food recognition task, such as AlexNet (Kagaya et al. 2014), GoogLeNet (Wu et al. 2016), Network-In-Networks (NIN) (Tanno et al. 2016), Inception V3 (Hassannejad et al. 2016), ResNet (Ming et al. 2018), and their combination (McAllister et al. 2018; Pandey et al. 2017). Recently, Martinel et al. (2018) combined extracted visual features from wide residual networks (WRNs) (Sergey and Nikos 2016) with ones from their proposed slice network for food recognition. To our knowledge, it achieves the state-of-the-art performance in benchmark datasets due to the high performance of WRNs.

In recipe-shared websites, food images are often associated with other rich content or context information, such as cuisines, ingredients, cooking methods, and food calories. Therefore, besides food recognition by the food type, food can be categorized by cuisines and other attributes, such as cuisine classification (Zhang 2011), taste, and flavor prediction (Druck 2013). What's more, different types of food labels, such as food name, food ingredients, and other attributes can also be learned simultaneously in a multi-task way. These tasks are very relevant and other tasks are generally helpful for visual feature learning to improve the performance of food recognition. For example, one or some of the following typical tasks, including recognizing food ingredients, classifying cooking methods, classifying restaurants, and predicting calorie value, are conducted simultaneously with food recognition (Chen and Ngo 2016; Ege and Yanai 2017; Min et al. 2017a; Zhang et al. 2016; Zhou and Lin 2016). One common way is joint food category and ingredient recognition. For example, Chen and Ngo (2016) developed different CNN architectures for multi-task learning for both food category and ingredient recognition. Zhou and Lin (2016) exploited rich ingredients and label relationships through bipartite-graph labels and then combined bipartite-graph labels and CNN together for both ingredient recognition and dish recognition. Recently, Aguilar et al. (2019) further proposed a new evaluation metric particularly for multi-task food analysis to simultaneously predict cuisine and food categories. There are also works (Min et al. 2017a; Wang et al. 2015) that fused features from different modalities including images and associated text for food recognition.

In addition, Kaur et al. (2017) augmented the deep neural network with noisy web food images to improve the performance of food recognition. Benefiting from large-scale food data from social media, some studies (Rich et al. 2016) (Barranco et al. 2016) learned to recognize food image content from social media, such as Instagram and Yelp.

5.2.2 Multiple-label Food Recognition and Detection. In real-world scenarios, there may be more than one food item in the image. The first work to recognize multiple-food items from one food image is proposed by Matsuda et al. (2012). They first detected candidate regions and then classified them. Matsuda and Yanai (2012) further exploited the co-occurrence relation information between food items for recognizing multiple-food meal photos. In addition, food detection and segmentation are widely used for images with multiple food items.

Food detection has earlier been considered as a binary classification problem, where the algorithm is used to distinguish whether one given image represents food or not, namely binary food detection (Kagaya et al. 2014; Ragusa et al. 2016). Both hand-crafted (Farinella et al. 2015a; Kitamura et al. 2009; Miyano et al. 2012) and deep features (Kagaya and Aizawa 2015; Meyers et al. 2015) are adopted. Compared with hand-crafted features, an improvement is achieved via CNN-based deep networks (Kagaya et al. 2014). CNN-based methods have been proposed for either feature extraction (Aguilar et al. 2017a; Ragusa et al. 2016) or the whole recognition process (Kagaya and Aizawa 2015; Singla et al. 2016). For example, Singla et al. (2016) used the GoogLeNet network for food/non-food classification. In addition for binary food detection, some works such as Bolaños and Radeva (2017) and Anzawa et al. (2019) used the deep network to recognize every food type present based on the detected regions. Different from food detection, food segmentation classifies each pixel from one food image. For example, recent research proposes an automatic weakly supervised method based on CNN (Shimoda and Yanai 2015) or distinct class-specific saliency maps (Shimoda and Yanai 2016). Besides food recognition by food items, there are some works on multi-label ingredient recognition (Bolaños et al. 2017; Chen et al. 2018). As one representative work, Aguilar et al. (2018) proposed a semantic food detection framework, which consists of three parts, namely food segmentation, food detection, and semantic food detection. Food segmentation uses the fully CNNs to produce the binary image and then adopts the Moore-Neighbor tracing algorithm to conduct boundary extraction. Food detection is achieved by retraining YOLOv2 (Redmon et al. 2016). Semantic food detection removes errors from food detection by combining results of segmentation and detection to obtain final food detection results.

5.2.3 Sensor-based Food Recognition and Monitoring. Over the past decade, a great variety of mobile devices and other sensors have been developed. Food recognition has been increasingly adapted into these sensors for health-aware applications. One general way is to apply food recognition to mobile devices. This also has other advantages of combined various built-in inertial sensors (Min et al. 2017c) with visual food recognition for monitoring activities of daily living, thus providing more complete information for dietary assessment and management (Kong and Tan 2011; Oliveira et al. 2014; Pouladzadeh et al. 2016a). For example, Kawano and Yanai (2015) proposed the mobile food recognition system FoodCam for calorie and nutrition estimation. Recently, deep learning–based mobile food recognition methods (Pouladzadeh and Shirmohammadi 2017; Tanno et al. 2016) have been developed. For example, Pouladzadeh and Shirmohammadi (2017) proposed a mobile recognition system that can recognize multiple food items in one meal, such as steak and potatoes, for further estimation on the nutrition and calories of the meal. However, when applying deep learning to mobile devices, some unique problems for mobile food recognition need to be solved, e.g., the complexity and memory requirements of deep learning solutions and energy consumption. Please refer to Ota et al. (2017) for more details in mobile deep learning. There are two types of mobile food recognition: client-server mode and client-mode. For the client-server mode, the mobile device is only used to take the picture and transfer it to the cloud, where food image processing is performed via the deep learning network (Merler et al. 2016; Peddi et al. 2017). For the client mode, food image processing is conducted in the mobile device. In this case, deep networks should be pruned or compressed to make them work in the mobile devices. For example, Yanai et al. (2016) compressed the deep network using product quantization for object recognition. They (Tanno et al. 2016) then used the compressed deep network for mobile food recognition. With the fast development of smart devices and food-related applications, we will witness more effective and efficient deep networks, such as MobileNets (Howard et al. 2017) and ShuffleNet (Zhang et al. 2017b) for mobile food recognition in the future.

There are also works on food recognition and monitoring in other sensors, such as acoustic-based, motion-based, and multimodal methods. For example, Yang et al. (2016) proposed an application iHearFood, which can use the Bluetooth headsets to analyze the chewing sound for food recognition via a deep network. Li et al. (2013) presented the design and implementation of a wearable oral sensory system to recognize human oral activities, such as chewing and drinking via sensing the teeth motion. Mirtchouk et al. (2016) used a multi-modal sensing device to combine the in-ear audio and head and wrist motions to more accurately classify the food type. A comprehensive survey on applying wearable sensors for automatic dietary monitoring is introduced in Schiboni and Amft (2018); refer to Schiboni and Amft (2018) for more details.

5.2.4 Food Portion Estimation. Estimating food portion size or food volume is necessary to estimate an individual's food and energy intake. Existing methods on image-based food portion estimation are divided into different types, including video-based or multiple images–based (Kong and Tan 2012; Mingui et al. 2010), two-images based (Dehais et al. 2017), and single-image-based ones (Fang et al. 2018; Meyers et al. 2015). Kong and Tan (2012) presented a mobile phone–based system, DietCam, which only requires users to take three images or a short video around the meal. Then three-dimensional (3D) models of visible food items will be reconstructed to estimate the volume of the food. Dehais et al. (2017) proposed a three-stage system to calculate portion sizes using only two images of a dish acquired by mobile devices with three stages. A dense 3D model is built from the two images to further serve to extract the volume of the different items. In contrast, Meyers et al. (2015) first modeled the correlation between RGB and depth image and then estimated the depth image from only one image. Finally, they used both RGB and estimated depth information for food volume estimation. Besides the CNNs, the generative adversarial networks are also used for food portion estimation (Fang et al. 2018). In addition, there are other calibration-based techniques for estimating food portion volume (Pouladzadeh et al. 2014). Although recent methods conducted food portion estimation from a single food image since this reduces a user's burden in the number and types of images that need to be acquired, accurate food portion estimation is still challenging due to large variations on food shapes and appearances.

5.2.5 Personalized Food Recognition. Personalized food image recognition focuses on classifying food images created for each individual user. It is very challenging due to dynamic datasets created by each user often have content with considerable variations between different users and a limited number of samples per person. There are few works in this area. Aizawa et al. (2013) conducted food image detection and food balance estimation using personal uploaded meal images. One recent work is Horiguchi et al. (2018), which adopted an incremental learning method to personalize a classifier for each user. Personalized food recognition will receive more attention because of its potentials in personalized food recommendation and multimedia foodlog.

In addition, there are some works on restaurant-specific food recognition. In the restaurant scenario, additional information such as location and menu information is utilized (Aguilar et al. 2018; Bettadapura et al. 2015; Herranz et al. 2017, 2015; Wang et al. 2016). For example, Xu et al. (2015) proposed a framework to incorporate geo-location information for dish classification. They trained the geolocalized models using these dish images with geographical locations, menus, and dish images. During the test stage, for one query, corresponding geolocalized models are selected and adapted to the query.

Tables 2 and 3 provide an overview of these approaches with respect to visual features, additional information, and recognition type. The classifiers that most methods adopt are SVM or Softmax. Table 4 shows an overview of current performance comparison on benchmark datasets.

5.2.6 Summary and Discussion. Food recognition has been widely studied in various fields, such as computer vision and multimedia. The key of food recognition is to extract discriminative visual features. Early researches on food recognition mainly extracted hand-crafted features. In the recent years, image recognition has undergone a paradigm shift towards using deep learning for its strong capability in feature learning, and food recognition is no exception. Compared with hand-crafted features, deep learning for food recognition has achieved great performance improvement. However, most of existing deep learning methods directly extracted deep visual features via CNNs and ignored characteristics of food images and are thus hard to achieve optimal performance. In contrast to general object recognition, food images typically do not exhibit distinctive spatial arrangement and common semantic patterns. One way to mitigate the problem is to utilize other rich content and context information from websites and social media. In addition, with the fast

Table 2. Summary of Food Recognition Using Conventional Visual Features

Reference	Visual Features	Additional Information	Recognition Type
(Bolle et al. 1996)	Texture, Color	-	Food recognition
(Puri et al. 2009)	Color, Textures	-	Mobile food recognition
(Wu and Yang 2009)	SIFT	-	Food recognition
(Joutou and Yanai 2010)	SIFT,Color, Texture	-	Food recognition
(Yang et al. 2010)	Pairwise Local Features Joint Pairwise Local Features	-	Food recognition
(Zong et al. 2010)	SIFT, Texture	-	Food recognition
(Bosch et al. 2011)	SIFT, Color, Texture	-	Food recognition
(Zhang 2011)	Color, Texture	-	Cuisine classification
(Matsuda and Yanai 2012)	SIFT, Color, HoG, Texture	-	Food recognition
(Matsuda et al. 2012)	SIFT, Color HoG, Texture	-	Food recognition
(Farinella et al. 2014b)	Texture	-	Food recognition
(Nguyen et al. 2014)	SIFT, Texture, Shape	-	Food recognition
(Anthimopoulos et al. 2014)	SIFT, Color	-	Food recognition
(Oliveira et al. 2014)	Color, Texture	-	Mobile food recognition
(Kawano and Yanai 2014c)	HoG, Color	-	Mobile food recognition
(Farinella et al. 2015a)	SIFT, Texture, Color	-	Food recognition
(Martinel et al. 2015)	Color, Shape, Texture	-	Food recognition
(Bettadapura et al. 2015)	SIFT, Color	Location & Menu	Restaurant-specific food recognition
(Farinella et al. 2015b)	SIFT, SPIN	-	Food recognition
(Kawano and Yanai 2015)	SIFT, Color, HoG	-	Mobile food recognition
(Ravl et al. 2015)	HoG, Texture, Color	-	Mobile food recognition
(Martinel et al. 2016)	SIFT, Color, Shape, Texture	-	Food recognition
(He et al. 2017)	Texture	-	Food recognition
(Zheng et al. 2017)	SIFT, Color	-	Food recognition

development of smart devices and sensing technologies, food recognition has been applied into mobile devices and other sensors for health-relevant applications. Consequently, new problems arise, e.g., the complexity and memory requirements of deep learning solutions and energy consumption when applying deep learning to mobile devices and other sensors, which is still a hot topic and needs further exploration.

5.3 Retrieval

These massive amounts of data shared on various sites allow gathering food-related data such as recipes, food images, and cooking videos. A food-relevant retrieval engine is necessary to obtain what we need. In real applications, the number of examples needed to train a food classifier may not be always available. In this case, food retrieval can be used to find similar foods among available ones and to suggest a possible food type. In health-oriented applications, predicting nutrition content and calorie information from food images requires fine-grained ingredient recognition. However, directly recognizing ingredients is sometimes challenging, since ingredients from prepared food are mixed and stirred. In this case, we can retrieve recipes based on the image query, namely cross-modal retrieval.

According to retrieval types, food-relevant retrieval consists of three types: visual food retrieval, recipe retrieval, and cross-modal recipe-image retrieval.

Table 3. Summary of Food Recognition Using Deep Visual Features

Reference	Visual Features	Additional Information	Recognition Type
(Kawano and Yanai 2014b)	HoG, Color, CNN	-	Food recognition
(Kagaya et al. 2014)	AlexNet	-	Food recognition
(Ao and Ling 2015)	GoogleNet	-	Food recognition
(Yanai and Kawano 2015)	AlexNet	-	Food recognition
(Christodoulidis et al. 2015)	CNN	-	Food recognition
(Wang et al. 2015)	VGG	Text	Recipe recognition
(Xu et al. 2015)	DeCAF	Location	Restaurant-specific food recognition
(Herranz et al. 2015)	DeCAF	Location	Restaurant-specific food recognition
(Herruzo et al. 2016)	GoogleNet	-	Food recognition
(Wang et al. 2016)	CNN	Location	Restaurant-specific food recognition
(Singla et al. 2016)	GoogleNet	-	Food recognition
(Ragusa et al. 2016)	AlexNet, VGG, NIN	-	Food recognition
(Wu et al. 2016)	GoogleNet	-	Food recognition
(Ciocca et al. 2016)	AlexNet	-	Food recognition
(Liu et al. 2016)	Inception	-	Food recognition
(Hassannejad et al. 2016)	Inception	-	Food recognition
(Tanno et al. 2016)	Network In Network	-	Mobile food recognition
(Chen and Ngo 2016)	VGG	Ingredients	Multi-task food recognition
(Zhang et al. 2016)	Designed network	Cooking method labels	Multi-task food recognition
(Wang et al. 2016)	Designed network	Restaurant labels	Multi-task food recognition
(Ege and Yanai 2017)	VGG	Food calories	Multi-task food recognition
(Min et al. 2017a)	DBM	Cuisine,Course	Multi-task cuisine recognition
(Aguilar et al. 2019)	VGG,ResNet	Cuisine,Dish	Multi-task food analysis
(Herranz et al. 2017)	AlexNet	Location & Menu	Restaurant-specific food recognition
(Bolaños and Radeva 2017)	GoogleNet	-	Food recognition
(Pandey et al. 2017)	AlexNet, GoogLeNet ResNet	-	Food recognition
(Chen et al. 2017e)	ResNet-152, DenseNet VGG-19	-	Food recognition
(Termritthikun et al. 2017)	NUInNet	-	Food recognition
(Kaur et al. 2017)	Inception-ResNet	-	Food recognition
(Pan et al. 2017)	AlexNet, CafffeNet RestNet-50	-	Ingredient classification
(Aguilar et al. 2017b)	InceptionV3, GoogLeNet ResNet-50	-	Food recognition
(McAllister et al. 2018)	ResNet-152, GoogleNet	-	Food recognition
(Ming et al. 2018)	ResNet-50	-	Mobile food recognition
(Martinel et al. 2018)	WISeR	-	Food recognition

Table 4. Performance Comparison on the Accuracy in Three Benchmark Datasets (%)

Reference	UECFood100	UECFood256	ETHZ Food-101
(Kawano and Yanai 2014b)	72.26	-	-
(Kawano and Yanai 2014c)	-	50.10	-
(Ravl et al. 2015)	53.35	-	-
(Martinel et al. 2015)	80.33	-	-
(Yanai and Kawano 2015)	78.77	67.57	70.41
(Ao and Ling 2015)	-	-	78.11
(Wu et al. 2016)	-	-	72.11
(Liu et al. 2016)	76.30	54.70	77.40
(Martinel et al. 2016)	84.31	-	55.89
(Hassannejad et al. 2016)	81.45	76.17	88.28
(Zheng et al. 2017)	70.84	-	-
(Bolaños and Radeva 2017)	-	63.16	79.20
(Aguilar et al. 2017b)	-	-	86.71
(Pandey et al. 2017)	-	-	72.12
(McAllister et al. 2018)	-	-	64.98
(Martinel et al. 2018)	**89.58**	**83.15**	**90.27**

For food image retrieval, image retrieval based on local descriptors (e.g., SIFT) has been extensively studied for over a decade due to their advantage in dealing with image transformations. For example, Kitamura et al. (2009) proposed a FoodLog system to retrieve personal food images via the combination of BoF visual features and SVM. Compared with Kitamura et al. (2009) and Aizawa et al. (2014) improved the food image retrieval system by supporting both image-based and text-based query. Some works, such as Farinella et al. (2016), further improved the performance of food image retrieval through the combination of different types of features, such as SIFT and Bag of Textons. Recently, image retrieval based on CNN has attracted increasing interest and demonstrated impressive performance. For example, Ciocca et al. (2018) adopted CNN-based features for food image retrieval, where different types of neural networks (e.g., VGG and ResNet) are used.

For recipe retrieval, the first step is generally to change the cooking instructions into structured representation for recipe representation. For example, Wang et al. (2008) modeled cooking instructions from Chinese recipes as graphs and further designed a novel similarity measurement to support efficient recipe searching. Recently, Chang et al. (2018) changed the recipe instruction into a tree-structure representation for recipe similarity calculation. In contrast, another type of method is to fuse different types of recipe-relevant features, such as cooking flow features, eating features, and nutrition features (Xie et al. 2011). Barlacchi et al. (2016) introduced a search engine for restaurant retrieval based on dishes one user wants to taste rather than using their general categories (such as Japanese and Italian). Finer-grained food properties, e.g., a particular way to cook a dish along with its specific ingredients are considered.

Besides food/recipe retrieval, different neural networks are designed to multimodal embedding for cross-modal recipe-image retrieval, such as attention network (Chen et al. 2017b) and multimodal deep Boltzmann machine (Min et al. 2017a). Another method for cross-modal retrieval is to use a hybrid neural network architecture, which jointly learned shared space via image and recipe embedding, where visual features are learned by CNN while recipe text features are sequentially modeled by Long-Short Term Memory (LSTM) (Carvalho et al. 2018; Salvador et al. 2017). As one representative work, Salvador et al. (2017) proposed a joint embedding model. There are mainly two components for a recipe, namely ingredients and cooking instructions. For ingredients, they

Table 5. Summary of Main Retrieval Methods

Reference	Data type		Dataset Name	Task
	Image	Text		
(Wang et al. 2008)	-	Cooking graph	Cooking graph database	Recipe retrieval
(Kitamura et al. 2009)	Food images	-	Foodlog	Food retrieval
(Xie et al. 2011)	-	Cooking graph	-	Recipe retrieval
(Barlacchi et al. 2016)	-	Dish name & Ingredients	Food Taste Knowledge Base (FKB)	Recipe retrieval
(Farinella et al. 2016)	Food images	-	UNICT-FD1200	Food retrieval
(Chen and Ngo 2016)	Food images	Ingredients	VIREO Food-172	Cross-modal retrieval
(Chen et al. 2017b)	Food images	Ingredients	-	Cross-modal retrieval
(Chen et al. 2017a)	Food images	Ingredients	-	Cross-modal retrieval
(Salvador et al. 2017)	Food images	Ingredients & Instructions	Recipe 1M	Cross-modal retrieval
(Min et al. 2017a)	Food images	Ingredients & Attributes	Yummly-28K	Cross-modal retrieval
(Ciocca et al. 2018)	Food images	-	Food524DB	Food retrieval
(Carvalho et al. 2018)	Food images	Ingredients & Instructions	Recipe 1M	Cross-modal retrieval

first extracted the ingredient name using bi-directional LSTM (Schuster and Paliwal 1997). Then each ingredient name is represented via the word2vec model (Mikolov et al. 2013). Finally, a bidirectional LSTM model is again used to encode these ingredients to the feature representation. For the cooking instruction, they utilized LSTM to encode it to a fixed-length feature representation. These two kinds of representations are concatenated to the final recipe representation. For the image representation, two deep convolutional networks, namely VGG-16 and Resnet-50 models, are adopted to extract visual features. Additional semantic regularization on the embedding is further introduced to improve joint embedding.

Table 5 provides a summary of main retrieval approaches with respect to features, dataset, and tasks.

Summary and Discussion. In this section, we identified three major types of food retrieval methods, namely food image retrieval, recipe retrieval, and cross-modal recipe-image retrieval. With the profusion of large-scale multimodal recipe collections, cross-modal recipe-image embedding and retrieval have received more attention. Different deep networks are proposed to solve this problem. Despite the progress made in cross-modal recipe retrieval, the retrieval performance is still very low. One key is incomplete food semantic understanding, because of its indistinctive spatial arrangement and irregular semantic patterns, leading to inaccurate correlation between food images and ingredients.

5.4 Recommendation

Food recommendation is an important domain for both individuals and society. Different from other types of recommendation systems, food recommendation involves more complex, multifaceted, and other context-dependent information (e.g., lifestyle preferences and culture) in predicting what people would like to eat. Taking all these factors into consideration, various recommendation methods are proposed and are divided into four types (Trattner and Elsweiler 2017a), namely collaborative filtering-based methods, content-based methods, hybrid methods, context-aware methods, and health-aware methods.

For content-based methods, recipe oriented recommendation has been extensively studied based on the similarity calculation between content items. For this type of food recommendation, different methods for recipe-based content representation are adopted, such as topic model–based representation (Kusmierczyk and Norvag 2016), structure-based representation (Jermsurawong and Habash 2015), and multi-modal representation with various attributes (Min et al. 2017b).

For collaborative filtering-based methods, classic singular value decomposition (Harvey et al. 2013) and matrix factorization (Ge et al. 2015) have been used widely to model the interaction between user and food items for recommendation. Other methods such as latent Dirichlet allocation and weighted matrix factorization are also used (Trattner and Elsweiler 2017b).

For context-aware approaches, numerous exploratory data analysis has demonstrated that rich context such as gender, time, hobbies, location, and cultural aspects is important in food recommendation. For example, Cheng et al. (2017) proposed to select users and items according to relevant context factors for context-aware food recommendation. In addition, exploring other factors such as culinary cultures can also be helpful for context-aware food recommendation. For example, Golder and Macy (2011) discovered some universal patterns regarding eating from millions of Twitter messages. Similar spatial-temporal patterns can be discovered by analyzing recipes, such as recipe preference distributions under different temporal intervals and regions (Wagner et al. 2014) (Kusmierczyk and Trattner 2015). For example, Silva et al. (2014) analyzed check-ins in Foursquare to identify the cultural differences and similarities across different geographical regions. Such cultural analysis and understanding from recipes and social media can help us develop recommendation mechanisms considering the cultural characterization of specific urban areas.

Health-aware food recommendation is unique. Incorporating health into the recommendation has largely been a recent focus (Markus et al. 2018; Nag et al. 2017b; Yang et al. 2017). Such a method not only caters to users' food preferences but should also be able to take users' health into account. For example, Nag et al. (2017b) proposed an online personalized nutrition recommendation system that can identify the healthiest items and recommend them to users based on their health data and environmental context. Recently, Markus et al. (2018) used different kinds of features from a recipe's title, ingredient list, cooking directions, popularity indicators (e.g., the number of ratings), and visual features to estimate the healthiness of recipes for health-aware recipe recommendation.

There are also works on mobile food recommendation (Maruyama et al. 2012) (Phanich et al. 2010). Other relevant studies in the field of nutrition science have shown that proper nutrition and health labels help people to make better food choice for food recommendation (Sonnenberg et al. 2013).

Summary and Discussion. Food recommendation has been becoming a hot research topic with many approaches proposed to improve the performance and experience of food recommendation from different aspects, such as incorporating rich context, multimodal learning, and introducing nutrition information. However, most existing methods mainly borrow ones from recommendation methods in other fields without considering characteristics of food recommendation, such as complex food preference. However, considering their great commercial potentials, we will look forward to seeing the surge, especially health-aware food recommendation in this research field.

5.5 Prediction and Monitoring

Online social media such as Twitter and Instagram provides its users with a way of recording their daily lives, such as dietary choices, leading to large-scale food data. They thus become rich sources to conduct food-related prediction and monitoring.

Many studies have adopted data-driven approaches to predict the income level (Ma et al. 2015), food consumption patterns (Mejova et al. 2015), recipe popularity (Sanjo and Katsurai 2017), and even diseases (Abbar et al. 2015) from these records in social media. Based on predicted results, using social media for monitoring public health will naturally be the next step (Capurro et al. 2014). For example, Sadilek et al. (2017) prevented the food-borne illness by mining the data in social media. They applied the machine learning method to Twitter data and developed a system that automatically detected venues likely to pose a public health hazard. Karisani and Agichtein (2018) detected personal health mentions in Twitter.

Summary and Discussion. We are living in the era of social media, leaving digital traces of various types of food-related activities online. Therefore, considering social media as one food social sensor, we can resort to social media for food-related prediction and monitoring, such as food consumption analysis and personal health mention prediction. However, it also presents researchers with some challenges, such as much noise and the sheer size of food data. Therefore, we expect more scalable data-driven methods for solving these problems in the future.

6 CHALLENGES

Food computing has received more attention in the last few years for its wide applications. Thus, it is extremely important to discuss existing challenges that form the major obstacles to current progress. This section presents key unresolved issues.

6.1 Food Image Recognition

Robust and accurate food image recognition is very essential for various health-oriented applications, such as food calorie estimation, food journaling, and automatic dietary management. However, it is very challenging for the following three reasons: (1) Food images have their own distinctive properties. They don't have any distinctive spatial layout. Although some food categories such as fruits, hamburgers, and pizzas have regular shapes, a large number of food dishes have deformable food appearance and are thus lack of rigid structures. Ingredients can be the constituent part of food. However, ingredients from multiple types of food images are distributed randomly on a plate. Other factors, such as cooking methods, also affect the appearance of food ingredients. This makes the task different from other ones like scene recognition, where we can always find some distinctive features such as buildings and trees. Therefore, simply borrowing the methods from object or scene recognition is hard to achieve satisfactory recognition results, especially for real-world applications, not to mention images with multiple-item meals. (2) Food image recognition belongs to fine-grained classification. Similarly, food image recognition encounters the same problem as the fine-grained classification, such as subtle differences among different food categories. However, we can not simply directly use existing fine-grained classification methods, such as (Fu et al. 2017) for food image recognition. The reason is that existing fine-grained categorization methods aim to distinguish between different breeds or species. They generally first discover the fixed semantic parts and then concatenate the features from both global object and semantic parts as the final representation. Such representation includes not only global features but also more discriminative local features. For example, in the bird classification, some semantic parts, such as head and breast, should first be localized. However, the concepts of common semantic parts do not exist in food images. Therefore, we should design a new fine-grained categorization paradigm, which is suitable for food recognition. (3) There is lack of large-scale benchmark food images with more categories. In the computer vision, the release of large-scale ImageNet dataset with the Wordnet ontology has greatly furthered the development of object recognition (Krizhevsky et al. 2012). Similarly, the large-scale food dataset is required. There are indeed some benchmark food datasets, such as Food101 (Bossard et al. 2014) and UEC Food256 (Kawano and Yanai 2014c). However, the categories and number of these datasets are not big enough compared with the ImageNet. In addition, food-oriented dataset construction has its particular challenges. For example, because of the region difference, there are probably several different names for the same dish. Similarly, some dishes are labeled with the same dish name but actually belong to different dishes with different ingredients. This means that it is harder to build a standard ontology according to the dish name like the Wordnet.

6.2 Vision-based Dietary Management System

With the fast development of computer vision and machine learning, more dietary management systems resort to vision-based methods. For example, Meyers et al. (2015) from Google proposed a system Im2Calories, which can recognize ingredients of the meal from one food image and then predict its calorie account. Beijbom et al. (2015) from Microsoft and University of California presented a computer vision system for automatically logging the food and calorie intake from food images in the restaurant scenario. However, existing dietary management systems are far from perfect and practical. The reasons derived are two-fold: (1) existing food recognition methods are robust to only few and standard dishes. In real-world scenarios, there are thousands of food categories to recognize. There are still considerable types of food images unavailable in the training set. As a result, the system fails to recognize the food, and then the estimated amount of calories is incorrect. In addition, most existing food recognition methods are not specifically for food images and thus have unsatisfactory recognition performance. (2) Even though we recognize the food and localize the food region, we next should estimate the food volume. It is still hard to accurately estimate the volume from one image. We can probably add the interaction to alleviate these problems, which conversely affect the user experience. Therefore, we should simultaneously solve the above-mentioned problems to enable a robust vision-based dietary management system, which is harder to achieve.

6.3 Multiple-network-oriented Food-data Fusion and Mining

During the past decade, the influence of social network services on people's daily life has sharply increased. Users participate in different social networks. For example, one user may share food photos in Instagram, upload the recipe to Twitter, and perform check-ins in Foursquare. To completely predict the health and wellness to deliver better healthcare, the first step is to effectively combine and integrate these food-related multi-modal signals from different social networks. However, the unbalanced data distributions in different networks and different accounts from different networks for each user make the effective fusion more challenging. Most food-relevant works mentioned previously use only one data source. They may not be enough to gain deeper insights and more complete knowledge from multi-source social media data. Furthermore, besides the social network, there are other types of networks, such as mobile networks and IoT. Therefore, we can obtain diverse signals from these different networks. For example, Fitocracy and MyFitnessPal provide the exercise semantics (i.e., sports activity type). Endomondo can be considered as a rich source of sequential data from wearable sensors and wellness-related ground truth. These mobile devices usually include rich multidimensional context information, such as altitude, longitude, latitude, and time. Computing users' lifestyles needs to be further integrated into these heterogeneous signals in a unified way. To the best of our knowledge, there are few publicly works towards it. Multimodal fusion faces still other challenges. For example, it is difficult to build one model that can exploit both shared and complementary information. In addition, not all the data sources will be helpful for certain food-related tasks in some cases. Among all these fused data sources, picking the useful ones is not an easy task.

6.4 Health-aware Personalized Food Recommendation

Existing methods (Elahi et al. 2017; Harvey et al. 2017) mainly refer to the trade-off for most users between recommending to the user what he/she wants and what is nutritionally appropriate, where the healthiness of the recipe can be predicted based on multiple cues, such as ingredients and images. However, there are other factors to make health-aware personalized food recommendation challenging, such as complex, multi-faceted information (e.g., the temporal and spatial context, culture, gender, and user preference). Each person is unique, and the physical state of each person is different at different moments. To enable more accurate food recommendation, we should monitor their wellness constantly. Although some works (Farseev and Chua 2017) integrated the data from wearable devices and several social networks to learn the wellness profile, the heterogeneous modality fusion is still difficult. Therefore, when developing health-aware personalized food recommendation systems, there are additional issues to consider that do not arise in other recommendation domains. These include that users may have various constrained needs, such as allergies or lifestyle preferences, and the desire to eat only fruit or vegetarian food. In such cases, existing methods do not work well.

6.5 Food Computing for Food Science

Food computing is an inherently multidisciplinary field, and its progress is predominantly dependent on support, knowledge, and advances in closely related fields, such as food science, biology, gastronomy, neuroscience, and computer science, as the performance of contemporary vision systems such as food image recognition is still far from perfect. Further investigations into the mechanisms of human perception on the visual food may be a crucially important step in gaining invaluable insights and relevant knowledge that can potentially inspire the better design of the dietary management. For example, most existing food computing methods mainly focus on the conventional multimodal data analysis and mining. However, food science involves multiple subdisciplines, such as food chemistry and food microbiology. We should cope with new data types (e.g., the chemical forms and the molecular structure in food) and new tasks (such as immunogenic epitopes detection from the wheat). Therefore, current food computing methods must be adapted or even re-designed to handle these new data and new tasks. For example, how to design a multimedia feature learning method to represent new data types, such as special chemical forms or the molecular structure in food? How to design novel food computing methods that target new tasks, such as ingredient recognition in the food engineering environment? How to use the food computing method to detect various food-borne illnesses in the food quantity control?

7 FUTURE DIRECTIONS

As mentioned earlier, considerable effort will be required in the future to tackle the challenges and open issues with food computing. Several future directions and solutions are listed as follows.

7.1 Large-scale Standard Food Dataset Construction

Like ImageNet for general objects in the computer vision, a large-scale ontology of ImageNet-level food images is also a critical resource for developing advanced, large-scale content-based food image search, classification, and understanding algorithms, as well as for providing critical training and benchmark data for such algorithms. To construct the large-scale food dataset, a feasible method is to combine food image crawling from social media and manual annotation from the crowd-sourcing platform AMT. In addition, we should consider the geographical distribution of food images, such as different cuisines, to cover the whole world. Each region has their own special cuisines and dishes; there is no food expert to master all the dishes. Therefore, the construction of the large-scale food dataset also should involve joint efforts of scientists all over the world.

7.2 Large-scale Robust Food Recognition System

Vision-based food system is very fundamental to various real-world applications, such as the dietary assessment and management system. The first priority is to develop a large-scale robust food recognition system. In recent years, deep learning approaches such as CNNs (Krizhevsky et al. 2012) and their variants (e.g., the VGG network (Szegedy et al. 2015), ResNet (He et al. 2016), and DenseNet (Huang et al. 2017)) have provided us with great opportunities to achieve this goal. Deep learning has the advantage of learning more abstract patterns progressively and automatically from raw image pixels in a multiple-layer architecture than using hand-engineered features. There are indeed some efforts for this direction. For example, Martinel et al. (2018) proposed a slice convolution network to capture vertical food structure and combined visual features from the general deep network to achieve the state-of-the art performance. We believe there are other special food structures and properties to explore. If we design the deep model to capture the structures particularly for food images from different aspects, the performance will be further improved. In addition, the constructed large-scale standard food dataset can also be critical to advance the development of food recognition system. There are more than 8K types of dishes worldwide, according to Wikipedia[8] (Bolaños et al. 2017). Compared with the large amount of dish types, the number of ingredients is limited. Therefore, one alternative solution is ingredient recognition. Some works (Bolaños et al. 2017; Chen and Ngo 2016) have conducted multi-label ingredient prediction from food images in terms of their lists of ingredients. Ingredient recognition will probably also be a solution for offering an automatic mechanism for recognize images for applications in easing the tracking of the nutrition habits, leading to more accurate dietary assessment.

7.3 Joint Deep and Broad Learning for Food Computing

A great amount of food-related data is being recorded in various modalities, such as text, images, and videos. It presents researchers with challenges, such as the sheer size of data, the difficulty in understanding recipes, computer vision, and other machine learning challenges to study the culinary culture, eating habits, and health. Fortunately, the recent breakthroughs in AI, especially the deep learning, provides powerful support for food data analysis from each data source. However, food-related entities are from different networks, such as social networks, recipe-sharing websites, and heterogeneous IoT sources. Effectively fusing these different information sources provides an opportunity for researchers to understand the food data more comprehensively, which makes "Broad Learning" an extremely important learning task. The aim of broad learning is to investigate principles, methodologies, and algorithms to discover synergistic knowledge across multiple data sources (Zhang et al. 2017a). Therefore, to learn, fuse, and mine multiple food-related information sources with large volumes and multi-modality, one future direction is to jointly combine deep learning and broad learning from different data sources into a unified multimedia food data fusion framework. Such a framework will provide a new paradigm, which is transformed to conventional food-related fields, such as food medicine and food science.

7.4 Food-oriented Multimodal Knowledge Graph Construction and Inference

We can exploit the enormous volume of food-related data using sophisticated data analysis techniques to discover patterns and new knowledge. However, to support heterogeneous modalities for more complex food-oriented retrieval, Question Answering (QA), reasoning, and inference,

[8] https://en.wikipedia.org/wiki/Lists_of_foods.

a more effective method is to build a food-oriented multimodal knowledge graph incorporating visual, textual, structured data, rich context information, as well as their diverse relations by learning from large-scale multimodal food data. In natural language processing, some promising results have been shown, e.g., Freebase (Bollacker et al. 2008). Semantic web technologies, e.g., ontologies and inference mechanism, have been used for the diabetes diet care (Li and Ko 2007). The study on visual relationships with triplets have been emerging in the area of computer vision, including the detection of visual relationships (Lu et al. 2016; Zhu and Jiang 2018) and generation of the scene graph (Johnson et al. 2015) from images. These technologies are helpful for constructing the visual web (Jain 2015). Other works, such as Zhu et al. (2015), tried to build a large-scale multimodal knowledge base system to support visual queries and has been shown as a promising way to construct the food-oriented multimodal knowledge graph. Such a multimodal knowledge graph is useful to consistently represent the food data from various heterogeneous data sources. In addition, the reasoning can also be conducted based on the knowledge graph for supporting complex query, QA, and multimodal dialog via the inference engines.

7.5 Food Computing for Personal Health

Modern multimedia research has been developed fast in some fields, such as art and entertainment, but lags in the health domain. Food is a fundamental element for the health. Food computing is emerging as a promising field for the health domain and can be used to quantify the lifestyle and navigate the personal health. Recently, some works, such as Nag et al. (2017a) and Nitish et al. (2017), have proposed the life navigation system for future health ecosystems, such as the cybernetic health. Karkar et al. (2017) proposed a TummyTrials app, which can aid a person in analyzing self-experiments to predict which type of food can trigger their symptoms. Food computing will provide principles and methodologies for the integration and understanding of food data produced by users. Combined with other information, such as attitudes and beliefs about food and recipes and the person's food preferences, lifestyles, and hobbies, we can construct the personal model for personalized and health-aware food recommendation service. Therefore, one important direction is to apply food computing to build the personal model for the health domain.

7.6 Food Computing for Human Behavior Understanding

Earlier studies have demonstrated that food affects human behavior (Kolata 1982). Different food choices lead to different changes in behaviors. For example, food additives and unhealthy diet could help to explain criminal behavior alcoholism.[9] There are also some works on the relationship between food and human behavior, such as the eating behavior (Achananuparp et al. 2018; Tsubakida et al. 2017), the brain activity (Rosenbaum et al. 2008b), and cooking activities (Damen et al. 2018; Stein and Mckenna 2013). For example, Achananuparp et al. (2018) used the data from MyFitnessPal to analyze healthy eating behaviors of users who actively record food diaries. Food computing can effectively utilize different food-oriented signals and thus will provide new methodologies and tools to advance the development in this direction.

7.7 Foodlog-oriented Food Computing

With the widespread use of mobile devices, e.g., digital cameras, smartphones, and tablets, people can easily take photos of food to record their diets. In addition, text-based meal record is also supported. Therefore, foodlogs record users' eating histories with multimodal signals. With the economic growth of the world, more people resort to foodlogs for recording their general diet via the smartphone. Foodlog-oriented food computing will become important for its multifarious

[9] https://articles.mercola.com/sites/articles/archive/2008/07/29/what-s-in-that-how-food-affects-your-behavior.aspx.

applications. (1) Foodlogs are most critical for health. Some works (Aizawa and Ogawa 2015; Kitamura et al. 2008; Waki et al. 2015) propose a food-logging system that is capable of distinguishing food images from other types of images for the analysis of food balance. For example, Aizawa and Ogawa (2015) have proposed the FoodLog system,[10] which can receive access to all sorts of dietary information based on photos sent by smartphones for the health management. To more precisely calculate daily intake of calories from these multimodal signals, a robust foodlog-oriented food recognition is also needed. (2) Foodlogs record what one eats or drinks daily and thus reflects their eating habits. Therefore, mining and analyzing rich foodlog data will enable personalized food recommendation that can offer healthier options for health-aware food recommendation (Trattner et al. 2017). In addition, foodlogs record current popular food. We can aggregate the foodlog data with time stamps from millions of uses for food popularity prediction.

7.8 Other Promising Applications in the Vertical Industry

There are other promising applications for food computing in vertical fields. For example, food computing can enable diverse applications in the smart-home field, such as smart kitchen and personal nutrition log. Smart-home systems can collect valuable information about users' preferences, nutrition intake, and health data via food computing methods, such as food recognition and cooking-video comprehension. Some existing works, such as Kojima et al. (2015), utilized the text information to understand the audio-visual scene for a cooking support robot. In the future, we believe that the smart kitchen robot needs more functions, more intelligent multimodal interaction, and dialog. Food recognition, recipe recommendation, and food-related text processing will work jointly to enable this goal. It will also play an important role in the smart farming. Existing works, such as Hernandez-Hernandez et al. (2017) and Chen et al. (2017c), can recognize and count the fruits in the trees. More food computing systems will be applied to help detect the illness of the food to guarantee the food safety and quantity. With the development of food computing, it will also be applied into more emerging vertical fields, such as smart retails (especially for the grocery shopping) and smart restaurants.

8 CONCLUSIONS

Food computing is a vibrant interdisciplinary field that aims to utilize computational approaches for acquiring and analyzing heterogeneous food data from disparate sources. With the increasing availability of large-scale food data, more food-oriented computational methods from different fields, such as computer vision and machine learning, will be widely used or quickly developed to enable the prosperity of food computing. Because of its interdisciplinary nature, it can be applied into many applications and services in various fields, from health, culture, agriculture, and medicine to biology. In this survey, we provide an extensive review of the most notable works to date on the datasets, definition, tasks, and applications of food computing. It is important to address future challenges based on the knowledge from past works and achievements.

Moving forward, the proposed food computing framework helps researchers understand current research and identify unresolved issues for future research. We also discuss some key challenges, particularly unique in food computing. For example, different from general object recognition, food does not always exhibit distinctive spatial layout and configuration. Therefore, a robust and accurate food image recognition is not trivial, and a new food recognition paradigm is vital to handle this. Current food recommendation is still in an initial stage of development and faces some challenges, such as dynamic and complex context modeling and accurate and robust food preference learning. Considering these challenges, some promising research directions are suggested, such

[10] http://www.foodlog.jp/en.

as large-scale standard food dataset construction, large-scale robust food recognition system, and food computing for foodlogs. These lines of promising directions need further research. Because of huge potentials in human health, culture, behavior, and other great commercial applications, we will look forward to seeing the surge in food computing in the future.

REFERENCES

Sofiane Abbar, Yelena Mejova, and Ingmar Weber. 2015. You tweet what you eat: Studying food consumption through Twitter. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. 3197–3206.

Palakorn Achananuparp, Ee Peng Lim, and Vibhanshu Abhishek. 2018. Does journaling encourage healthier choices? Analyzing healthy eating behaviors of food journalers. In *Proceedings of the International Digital Health Conference*.

Eduardo Aguilar, Marc Bolaños, and Petia Radeva. 2017a. Exploring food detection using CNNs. In *Proceedings of the International Conference on Computer Aided Systems Theory*. 339–347.

Eduardo Aguilar, Marc Bolaños, and Petia Radeva. 2017b. Food recognition using fusion of classifiers based on CNNs. In *Proceedings of the International Conference on Image Analysis and Processing*. 213–224.

Eduardo Aguilar, Marc Bolaños, and Petia Radeva. 2019. Regularized uncertainty-based multi-task learning model for food analysis. *J. Vis. Comm. Image Rep.* 60 (2019), 360–370.

E. Aguilar, B. Remeseiro, M. Bolaños, and P. Radeva. 2018. Grab, pay and eat: Semantic food detection for smart restaurants. *IEEE Trans. Multimed.* 20, 12 (2018), 3266–3275.

Yong-Yeol Ahn, Sebastian E. Ahnert, James P. Bagrow, and Albert-László Barabási. 2011. Flavor network and the principles of food pairing. *Sci. Rep.* 1, 7377 (2011), 196.

K. Aizawa, K. Maeda, M. Ogawa, Y. Sato, M. Kasamatsu, K. Waki, and H. Takimoto. 2014. Comparative study of the routine daily usability of FoodLog: A smartphone-based food recording tool assisted by image retrieval. *J. Diab. Sci. Technol.* 8, 2 (2014), 203–208.

Kiyoharu Aizawa, Yuto Maruyama, He Li, and Chamin Morikawa. 2013. Food balance estimation by using personal dietary tendencies in a multimedia food log. *IEEE Trans. Multimed.* 15, 8 (2013), 2176–2185.

Kiyoharu Aizawa and Makoto Ogawa. 2015. FoodLog: Multimedia tool for healthcare applications. *IEEE MultiMedia* 22, 2 (2015), 4–8.

M. M. Anthimopoulos, L. Gianola, L. Scarnato, P. Diem, and S. G. Mougiakakou. 2014. A food recognition system for diabetic patients based on an optimized bag-of-features model. *IEEE J. Biomed. Health Inform.* 18, 4 (2014), 1261–1271.

Masashi Anzawa, Sosuke Amano, Yoko Yamakata, Keiko Motonaga, Akiko Kamei, and Kiyoharu Aizawa. 2019. Recognition of multiple food items in a single photo for use in a buffet-style restaurant. *IEICE Transactions* 102-D, 2 (2019), 410–414.

Shuang Ao and Charles X. Ling. 2015. Adapting new categories for food recognition with deep representation. In *Proceedings of the IEEE International Conference on Data Mining Workshop*. 1196–1203.

Gianni Barlacchi, Azad Abad, Emanuele Rossinelli, and Alessandro Moschitti. 2016. Appetitoso: A search engine for restaurant retrieval based on dishes. In *Proceedings of the Third Italian Conference on Computational Linguistics CLiC-it*. 46–50.

Roberto Camacho Barranco, Laura M. Rodriguez, Rebecca Urbina, and M. Shahriar Hossain. 2016. Is a picture worth ten thousand words in a review dataset? Retrieved from: *arXiv:1606.07496* (2016).

Oscar Beijbom, Neel Joshi, Dan Morris, Scott Saponas, and Siddharth Khullar. 2015. Menu-match: Restaurant-specific food logging from images. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*. 844–851.

D. Bell. 1997. Consuming geographies: We are where we eat. Routledge.

Vinay Bettadapura, Edison Thomaz, Aman Parnami, Gregory D. Abowd, and Irfan Essa. 2015. Leveraging context to support automated food recognition in restaurants. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*. 580–587.

Marc Bolaños, Aina Ferrà, and Petia Radeva. 2017. Food ingredients recognition through multi-label learning. In *International Conference on Image Analysis and Processing*. 394–402.

Marc Bolaños and Petia Radeva. 2017. Simultaneous food localization and recognition. In *Proceedings of the International Conference on Pattern Recognition*. 3140–3145.

Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: A collaboratively created graph database for structuring human knowledge. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*. 1247–1250.

Ruud M. Bolle, Jonathan H. Connell, Norman Haas, Rakesh Mohan, and Gabriel Taubin. 1996. VeggieVision: A produce recognition system. In *Proceedings of the IEEE Workshop on Applications of Computer Vision*. 244–251.

M. Bosch, F. Zhu, N. Khanna, C. J. Boushey, and E. J. Delp. 2011. Combining global and local features for food identification in dietary assessments. In *Proceedings of the IEEE International Conference on Image Processing*. 1789–1792.

Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. 2014. Food-101-mining discriminative components with random forests. In *Proceedings of the European Conference on Computer Vision*. 446–461.

Tamara Bucher, Klazine Van der Horst, and Michael Siegrist. 2013. Fruit for dessert. How people compose healthier meals. *Appetite* 60 (2013), 74–80.

B. V. R. Silva and J. Cui. 2017. A survey on automated food monitoring and dietary management systems. *J. Health Med. Inform.* 8, 3 (2017).

L. Canetti, E. Bachar, and E. M. Berry. 2002. Food and emotion. *Behav. Proc.* 60, 2 (2002), 157–164.

Daniel Capurro, Kate Cole, Maria I. Echavarría, Jonathan Joe, Tina Neogi, and Anne M. Turner. 2014. The use of social networking sites for public health practice and research: A systematic review. *J. Med. Internet Res.* 16, 3 (2014), e79.

Micael Carvalho, Rémi Cadène, David Picard, Laure Soulier, Nicolas Thome, and Matthieu Cord. 2018. Cross-modal retrieval in the cooking context: Learning semantic text-image embeddings. In *Proceedings of the 41st International ACM SIGIR Conference on Research Development in Information Retrieval.* 35–44.

Minsuk Chang, Leonore V. Guillain, Hyeungshik Jung, Vivian M. Hare, Juho Kim, and Maneesh Agrawala. 2018. RecipeScape: An interactive tool for analyzing cooking instructions at scale. In *Proceedings of the Conference on Human Factors in Computing Systems.* 451:1–451:12.

Itthi Chatnuntawech, Kittipong Tantisantisom, Paisan Khanchaitit, Thitikorn Boonkoom, Berkin Bilgic, and Ekapol Chuangsuwanich. 2018. Rice classification using spatio-spectral deep convolutional neural network. Retrieved from: *arXiv:1805.11491* (2018).

Hao Chen, Jianglong Xu, Guangyi Xiao, Qi Wu, Shiqin Zhang, Hao Chen, Jianglong Xu, Guangyi Xiao, Qi Wu, and Shiqin Zhang. 2018. Fast auto-clean CNN model for online prediction of food materials. *J. Parallel Distrib. Comput.* 117 (2018), 218–227.

Jingjing Chen, Chong-Wah Ngo, and Tat-Seng Chua. 2017a. Cross-modal recipe retrieval with rich food attributes. In *Proceedings of the ACM on Multimedia Conference.* 1771–1779.

Jingjing Chen and Chong-Wah Ngo. 2016. Deep-based ingredient recognition for cooking recipe retrieval. In *Proceedings of the ACM on Multimedia Conference.* 32–41.

Jingjing Chen, Lei Pang, and Chong Wah Ngo. 2017b. Cross-modal recipe retrieval: How to cook this dish? In *Proceedings of the International Conference on Multimedia Modeling.* 588–600.

Mei Chen, Kapil Dhingra, Wen Wu, Lei Yang, Rahul Sukthankar, and Jie Yang. 2009. PFID: Pittsburgh fast-food image dataset. In *Proceedings of the IEEE International Conference on Image Processing.* 289–292.

Mei Yun Chen, Yung Hsiang Yang, Chia Ju Ho, Shih Han Wang, Shane Ming Liu, Eugene Chang, Che Hua Yeh, and Ouhyoung Ming. 2012. Automatic Chinese food identification and quantity estimation. In *Proceedings of the SIGGRAPH Asia 2012 Technical Briefs.* 29.

Steven W. Chen, Shreyas S. Skandan, Sandeep Dcunha, Jnaneshwar Das, Edidiong Okon, Chao Qu, Camillo Jose Taylor, and Vijay Kumar. 2017c. Counting apples and oranges with deep learning: A data driven approach. *IEEE Robot. Automat. Lett.* 2, 2 (2017), 781–788.

Xin Chen, Hua Zhou, and Liang Diao. 2017e. ChineseFoodNet: A large-scale image dataset for Chinese food recognition. Retrieved from: *CoRR* abs/1705.02743 (2017).

Zikuan Chen and Yang Tao. 2001. Food safety inspection using "from presence to classification" object-detection model. *Pattern Recog.* 34, 12 (2001), 2331–2338.

Hao Cheng, Markus Rokicki, and Eelco Herder. 2017. The influence of city size on dietary choices and food recommendation. In *Proceedings of the Conference on User Modeling, Adaptation and Personalization.* 359–360.

Stergios Christodoulidis, Marios Anthimopoulos, and Stavroula Mougiakakou. 2015. Food recognition for dietary assessment using deep convolutional neural networks. In *Proceedings of the International Conference on Image Analysis and Processing.* 458–465.

J. Chung, J. Chung, W. Oh, Y. Yoo, W. G. Lee, and H. Bang. 2017. A glasses-type wearable device for monitoring the patterns of food intake and facial activity. *Sci. Rep.* 7 (2017), 41690.

Gianluigi Ciocca, Paolo Napoletano, and Raimondo Schettini. 2015. Food recognition and leftover estimation for daily diet monitoring. In *Proceedings of the International Conference on Image Analysis and Processing.* 334–341.

Gianluigi Ciocca, Paolo Napoletano, and Raimondo Schettini. 2016. Food recognition: A new dataset, experiments, and results. *IEEE J. Biomed. Health Inform.* 21, 3 (2016), 588–598.

Gianluigi Ciocca, Paolo Napoletano, and Raimondo Schettini. 2017. Learning CNN-based features for retrieval of food images. In *Proceedings of the International Conference on Image Analysis and Processing.* 426–434.

Gianluigi Ciocca, Paolo Napoletano, and Raimondo Schettini. 2018. CNN-based features for retrieval and classification of food images. *Comput. Vis. Image Underst.* 176–177 (2018), 70–77.

Felicia Cordeiro, Elizabeth Bales, Erin Cherry, and James Fogarty. 2015a. Rethinking the mobile food journal: Exploring opportunities for lightweight photo-based capture. In *Proceedings of the ACM Conference on Human Factors in Computing Systems.* 3207–3216.

F. Cordeiro, D. A. Epstein, E. Thomaz, E. Bales, A. K. Jagannathan, G. D. Abowd, and J. Fogarty. 2015b. Barriers and negative nudges: Exploring challenges in food journaling. In *Proceedings of the ACM Conference on Human Factors in Computing Systems.* 1159–1162.

Aron Culotta. 2014. Estimating county health statistics with Twitter. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. 1335–1344.

Dima Damen, Hazel Doughty, Giovanni Maria Farinella, Sanja Fidler, Antonino Furnari, Evangelos Kazakos, Davide Moltisanti, Jonathan Munro, Toby Perrett, Will Price, and Michael Wray. 2018. Scaling egocentric vision: The EPIC-KITCHENS dataset. In *Proceedings of the European Conference on Computer Vision*. 753–771.

Joachim Dehais, Marios Anthimopoulos, Sergey Shevchik, and Stavroula Mougiakakou. 2017. Two-view 3D reconstruction for food volume estimation. *IEEE Trans. Multimed.* 19, 5 (2017), 1090–1099.

Gregory Druck. 2013. Recipe attribute prediction using review text as supervision. In *Proceedings of the International Joint Conference on Artificial Intelligence Workshop on Cooking with Computers*.

Takumi Ege and Keiji Yanai. 2017. Simultaneous estimation of food categories and calories with multi-task CNN. In *Proceedings of the 15th IAPR International Conference on Machine Vision Applications*. 198–201.

Mehdi Elahi, David Elsweiler, Georg Groh, Morgan Harvey, Bernd Ludwig, Francesco Ricci, and Alan Said. 2017. User nutrition modelling and recommendation: Balancing simplicity and complexity. In *Proceedings of the Conference on User Modeling, Adaptation and Personalization*. 93–96.

David Elsweiler, Morgan Harvey, and Morgan Harvey. 2017. Exploiting food choice biases for healthier recipe recommendation. In *International ACM SIGIR Conference on Research and Development in Information Retrieval*. 575–584.

Shaobo Fang, Zeman Shao, Runyu Mao, Chichen Fu, Deborah A. Kerr, Carol J. Boushey, Edward J. Delp, and Fengqing Zhu. 2018. Single-view food portion estimation: Learning image-to-energy mappings using generative adversarial networks. In *Proceedings of the IEEE International Conference on Image Processing*. 251–255.

G. M. Farinella, D. Allegra, M. Moltisanti, F. Stanco, and S. Battiato. 2016. Retrieval and classification of food images. *Comput. Biol. Med.* 77 (2016), 23–39.

Giovanni Maria Farinella, Dario Allegra, and Filippo Stanco. 2014a. A benchmark dataset to study the representation of food images. In *Proceedings of the European Conference on Computer Vision*. 584–599.

Giovanni Maria Farinella, Dario Allegra, Filippo Stanco, and Sebastiano Battiato. 2015a. On the exploitation of one class classification to distinguish food vs non-food images. In *Proceedings of the International Conference on Image Analysis and Processing*. 375–383.

Giovanni Maria Farinella, Marco Moltisanti, and Sebastiano Battiato. 2014b. Classifying food images represented as bag of textons. In *Proceedings of the IEEE International Conference on Image Processing*. 5212–5216.

Giovanni Maria Farinella, Marco Moltisanti, and Sebastiano Battiato. 2015b. Food recognition using consensus vocabularies. In *Proceedings of the International Conference on Image Analysis and Processing*. 384–392.

Aleksandr Farseev and Tat-Seng Chua. 2017. Tweet can be fit: Integrating data from wearable sensors and multiple social networks for wellness profile learning. *ACM Trans. Inform. Syst.* 35, 4 (2017), 42:1–42:34.

Jianlong Fu, Heliang Zheng, and Tao Mei. 2017. Look closer to see better: Recurrent attention convolutional neural network for fine-grained image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4476–4484.

Mouzhi Ge, Mehdi Elahi, Ignacio Fernández-Tobías, Francesco Ricci, and David Massimo. 2015. Using tags and latent factors in a food recommender system. In *Proceedings of the International Conference on Digital Health*. 105–112.

Andrea Giampiccoli and Janet Hayward Kalis. 2012. Tourism, food, and culture: Community-based tourism, local food, and community development in Mpondoland. *Cult. Agricult.* 34, 2 (2012), 101–123.

S. A. Golder and M. W. Macy. 2011. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science* 333, 6051 (2011), 1878–1881.

C. M. Hall and C. M. Hall. 2003. *Wine, Food, and Tourism Marketing*. Routledge.

Jun Harashima, Yuichiro Someya, and Yohei Kikuta. 2017. Cookpad image dataset: An image collection as infrastructure for food research. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1229–1232.

Caleb Harper and Mario Siller. 2015. OpenAG: A globally distributed network of food computing. *IEEE Pervas. Comput.* 14, 4 (2015), 24–27.

Marvin Harris. 1985. Good to eat: Riddles of food and culture. *Amer. Anthrop.* 2 (1985).

Morgan Harvey, Bernd Ludwig, and David Elsweiler. 2013. You are what you eat: Learning user tastes for rating prediction. In *Proceedings of the International Symposium on String Processing and Information Retrieval*. 153–164.

Hamid Hassannejad, Guido Matrella, Paolo Ciampolini, Ilaria De Munari, Monica Mordonini, and Stefano Cagnoni. 2016. Food image recognition using very deep convolutional networks. In *Proceedings of the International Workshop on Multimedia Assisted Dietary Management*. 41–49.

Hongsheng He, Fanyu Kong, and Jindong Tan. 2017. DietCam: Multiview food recognition using a multikernel SVM. *IEEE J. Biomed. Health Inform.* 20, 3 (2017), 848–855.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 770–778.

Jose Luis Hernandez-Hernandez, Mario Hernandez-Hernandez, Severino Feliciano-Morales, Valentín Álvarez-Hilario, and Israel Herrera-Miranda. 2017. Search for optimum color space for the recognition of oranges in agricultural fields. In *Proceedings of the International Conference on Technologies and Innovation.* 296–307.

Luis Herranz, Shuqiang Jiang, and Ruihan Xu. 2017. Modeling restaurant context for food recognition. *IEEE Trans. Multimed.* 19, 2 (2017), 430–440.

Luis Herranz, Ruihan Xu, and Shuqiang Jiang. 2015. A probabilistic model for food image recognition in restaurants. In *Proceedings of the IEEE International Conference on Multimedia and Expo.* 1–6.

P. Herruzo, M.Bolaños, and P. Radeva. 2016. Can a CNN recognize Catalan diet? In *Proceedings of the American Institute of Physics Conference Series.* 211–252.

Hajime Hoashi, Taichi Joutou, and Keiji Yanai. 2010. Image recognition of 85 food categories by feature fusion. In *Proceedings of the IEEE International Symposium on Multimedia.* 296–301.

S. Horiguchi, S. Amano, M. Ogawa, and K. Aizawa. 2018. Personalized classifier for food image recognition. *IEEE Trans. Multimed.* 20, 10 (2018), 2836–2848.

Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. 2017. MobileNets: Efficient convolutional neural networks for mobile vision applications. Retrieved from: *CoRR* abs/1704.04861 (2017).

G. Huang, Z. Liu, L. v. d. Maaten, and K. Q. Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 2261–2269.

Ramesh Jain. 2015. Let's weave the visual web. *IEEE Multimedia* 22, 3 (2015), 66–72.

Jermsak Jermsurawong and Nizar Habash. 2015. Predicting the structure of cooking recipes. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing.* 781–786.

A. R. Jimenez, A. K. Jain, R. Ceres, and J. L. Pons. 1999. Automatic fruit recognition: A survey and new results using range/attenuation images. *Pattern Recog.* 32, 10 (1999), 1719–1736.

Justin Johnson, Ranjay Krishna, Michael Stark, Li Jia Li, David A. Shamma, Michael S. Bernstein, and Fei Fei Li. 2015. Image retrieval using scene graphs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 3668–3678.

Taichi Joutou and Keiji Yanai. 2010. A food image recognition system with multiple kernel learning. In *Proceedings of the IEEE International Conference on Image Processing.* 285–288.

Hokuto Kagaya and Kiyoharu Aizawa. 2015. Highly accurate food/non-food image classification based on a deep convolutional neural network. In *Proceedings of the International Conference on Image Analysis and Processing.* 350–357.

Hokuto Kagaya, Kiyoharu Aizawa, and Makoto Ogawa. 2014. Food detection and recognition using convolutional neural network. In *Proceedings of the ACM International Conference on Multimedia.* 1085–1088.

Payam Karisani and Eugene Agichtein. 2018. Did you really just have a heart attack? Towards robust detection of personal health mentions in social media. Retrieved from: *CoRR* abs/1802.09130 (2018).

Ravi Karkar, Jessica Schroeder, Daniel A. Epstein, Laura R. Pina, Jeffrey Scofield, James Fogarty, Julie A. Kientz, Sean A. Munson, Roger Vilardaga, and Jasmine Zia. 2017. TummyTrials: A feasibility study of using self-experimentation to detect individualized food triggers. In *Proceedings of the Conference on Human Factors in Computing Systems.* 6850–6863.

Parneet Kaur, Karan Sikka, and Ajay Divakaran. 2017. Combining weakly and webly supervised learning for classifying food images. Retrieved from: abs/1712.08730 (2017).

Yoshiyuki Kawano and Keiji Yanai. 2014a. Automatic expansion of a food image dataset leveraging existing categories with domain adaptation. In *Proceedings of the European Conference on Computer Vision.* 3–17.

Yoshiyuki Kawano and Keiji Yanai. 2014b. Food image recognition with deep convolutional features. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing.* 589–593.

Yoshiyuki Kawano and Keiji Yanai. 2014c. Foodcam-256: A large-scale real-time mobile food recognition system employing high-dimensional features and compression of classifier weights. In *Proceedings of the ACM International Conference on Multimedia.* 761–762.

Yoshiyuki Kawano and Keiji Yanai. 2015. FoodCam: A real-time food recognition system on a smartphone. *Multimed. Tools Applic.* 74, 14 (2015), 5263–5287.

Sunil K. Khanna. 2009. Food and culture: A reader (2nd Ed.). *Ecol. Food Nutrit.* 48, 2 (2009), 157–159.

W. D. Killgore, A. D. Young, L. A. Femia, P. Bogorodzki, J. Rogowska, and D. A. Yurgeluntodd. 2003. Cortical and limbic activation during viewing of high- versus low-calorie foods. *Neuroimage* 19, 4 (2003), 1381–1394.

W. D. Killgore and D. A. Yurgelun-Todd. 2005. Body mass predicts orbitofrontal activity during visual presentations of high-calorie foods. *Neuroreport* 16, 8 (2005), 859–863.

Kyung Joong Kim and Chang Ho Chung. 2016. Tell me what you eat, and I will tell you where you come from: A data science approach for global recipe data on the web. *IEEE Access* 4 (2016), 8199–8211.

Osame Kinouchi, Rosa W. Diezgarcia, Adriano J. Holanda, Pedro Zambianchi, and Antonio C. Roque. 2008. The nonequilibrium nature of culinary evolution. *New J. Phys.* 10, 7 (2008), 073020.

Keigo Kitamura, Toshihiko Yamasaki, and Kiyoharu Aizawa. 2008. Food log by analyzing food images. In *Proceedings of the ACM International Conference on Multimedia*. 999–1000.

Keigo Kitamura, Toshihiko Yamasaki, and Kiyoharu Aizawa. 2009. FoodLog: Capture, analysis and retrieval of personal food images via web. In *Proceedings of the ACM Multimedia Workshop on Multimedia for Cooking and Eating Activities*. 23–30.

Simon Knez and Luka Šajn. 2015. Food object recognition using a mobile device: State of the art. In *Proceedings of the International Conference on Image Analysis and Processing*. 366–374.

Ryosuke Kojima, Osamu Sugiyama, and Kazuhiro Nakadai. 2015. Audio-visual scene understanding utilizing text information for a cooking support robot. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. 4210–4215.

G. Kolata. 1982. Food affects human behavior. *Science* 218, 4578 (1982), 1209–10.

Fanyu Kong and Jindong Tan. 2011. DietCam: Regular shape food recognition with a camera phone. In *Proceedings of the International Conference on Body Sensor Networks*. 127–132.

Fanyu Kong and Jindong Tan. 2012. DietCam: Automatic dietary assessment with mobile camera phones. *Pervas. Mobile Comput.* 8, 1 (2012), 147–163.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. In *Proceedings of the International Conference on Neural Information Processing Systems*. 1097–1105.

H. Kuehne, A. Arslan, and T. Serre. 2014. The language of actions: Recovering the syntax and semantics of goal-directed human activities. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 780–787.

Tomasz Kusmierczyk and Kjetil Norvag. 2016. Online food recipe title semantics: Combining nutrient facts and topics. In *Proceedings of the ACM International on Conference on Information and Knowledge Management*. 2013–2016.

Tomasz Kusmierczyk and Christoph Trattner. 2015. Temporal patterns in online food innovation. In *Proceedings of the International Conference on World Wide Web*. 1345–1350.

Yann LeCun, Yoshua Bengio, and Geoffrey E. Hinton. 2015. Deep learning. *Nature* 521, 7553 (2015), 436–444.

Cheng-Yuan Li, Yen-Chang Chen, Wei-Ju Chen, Polly Huang, and Hao-hua Chu. 2013. Sensor-embedded teeth for oral activity recognition. In *Proceedings of the International Symposium on Wearable Computers*. 41–44.

Huan Chung Li and Wei Min Ko. 2007. Automated food ontology construction mechanism for diabetes diet care. In *Proceedings of the International Conference on Machine Learning and Cybernetics*. 2953–2958.

Yanchao Liang and Jianhua Li. 2017. Computer vision-based food calorie estimation: Dataset, method, and experiment. Retrieved from: *arXiv preprint arXiv:1705.07632* (2017).

Chang Liu, Yu Cao, Yan Luo, Guanling Chen, Vinod Vokkarane, and Yunsheng Ma. 2016. Deepfood: Deep learning-based food image recognition for computer-aided dietary assessment. In *Proceedings of the International Conference on Smart Homes and Health Telematics*. 37–48.

David G. Lowe. 2004. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* 60, 2 (2004), 91–110.

Cewu Lu, Ranjay Krishna, Michael Bernstein, and Li Fei-Fei. 2016. Visual relationship detection with language priors. In *Proceedings of the European Conference on Computer Vision*. 852–869.

Yuzhen Lu, Yuping Huang, and Renfu Lu. 2017. Innovative hyperspectral imaging-based techniques for quality evaluation of fruits and vegetables: A review. *Appl. Sci.* 7, 2 (2017), 189.

Chao Ma, Ze Song, Xuhui Sun, and Guangchuan Zhao. 2015. Will low-income populations love spicy foods more? Accounting for tastes. University Library of Munich, Germany.

Rokicki Markus, Trattner Christoph, and Herder Eelco. 2018. The impact of recipe features, social cues and demographics on estimating the healthiness of online recipes. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*.

C. K. Martin, J. B. Correa, H. Han, H. R. Allen, J. C. Rood, C. M. Champagne, B. K. Gunturk, and G. A. Bray. 2012. Validity of the remote food photography method (RFPM) for estimating energy and nutrient intake in near real-time. *Obesity* 20, 4 (2012), 891–899.

Niki Martinel, Gian Luca Foresti, and Christian Micheloni. 2018. Wide-slice residual networks for food recognition. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*. 567–576.

Niki Martinel, Claudio Piciarelli, and Christian Micheloni. 2016. A supervised extreme learning committee for food recognition. *Comput. Vis. Image Underst.* 148 (2016), 67–86.

Niki Martinel, Claudio Piciarelli, Christian Micheloni, and Gian Luca Foresti. 2015. A structured committee for food recognition. In *Proceedings of the IEEE International Conference on Computer Vision Workshop*. 484–492.

Takuma Maruyama, Yoshiyuki Kawano, and Keiji Yanai. 2012. Real-time mobile recipe recommendation system using food ingredient recognition. In *Proceedings of the ACM International Workshop on Interactive Multimedia on Mobile and Portable Devices*. 27–34.

Yuji Matsuda, Hajime Hoashi, and Keiji Yanai. 2012. Recognition of multiple-food images by detecting candidate regions. In *Proceedings of the IEEE International Conference on Multimedia and Expo*. 25–30.

Y. Matsuda and K. Yanai. 2012. Multiple-food recognition considering co-occurrence employing manifold ranking. In *Proceedings of the International Conference on Pattern Recognition.* 2017–2020.

Patrick McAllister, Huiru Zheng, Raymond Bond, and Anne Moorhead. 2018. Combining deep residual neural network features with supervised machine learning algorithms to classify diverse food image datasets. *Comput. Biol. Med.* 95 (2018), 217–233.

K. McCrickerd and C. G. Forde. 2016. Sensory influences on food intake control: Moving beyond palatability. *Obesity Reviews* 17, 1 (2016), 18–29.

Yelena Mejova, Sofiane Abbar, and Hamed Haddadi. 2016. Fetishizing food in digital age: #Foodporn around the world. In *Proceedings of the International Conference on Weblogs and Social Media.* 250–258.

Yelena Mejova, Hamed Haddadi, Anastasios Noulas, and Ingmar Weber. 2015. # FoodPorn: Obesity patterns in culinary interactions. In *Proceedings of the International Conference on Digital Health.* 51–58.

Michele Merler, Hui Wu, Rosario Uceda-Sosa, Quoc-Bao Nguyen, and John R. Smith. 2016. Snap, Eat, RepEat: A food recognition engine for dietary logging. In *Proceedings of the International Workshop on Multimedia Assisted Dietary Management.* 31–40.

Austin Meyers, Nick Johnston, Vivek Rathod, Anoop Korattikara, Alex Gorban, Nathan Silberman, Sergio Guadarrama, George Papandreou, Jonathan Huang, and Kevin P. Murphy. 2015. Im2Calories: Towards an automated mobile vision food diary. In *Proceedings of the IEEE International Conference on Computer Vision.* 1233–1241.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *CoRR* arXiv:1905.04097 (2013).

Weiqing Min, Bing-Kun Bao, Shuhuan Mei, Yaohui Zhu, Yong Rui, and Shuqiang Jiang. 2018. You are what you eat: Exploring rich recipe information for cross-region food analysis. *IEEE Trans. Multimed.* 20, 4 (2018), 950–964.

Weiqing Min, Shuqiang Jiang, Jitao Sang, Huayang Wang, Xinda Liu, and Luis Herranz. 2017a. Being a supercook: Joint food attributes and multi-modal content modeling for recipe retrieval and exploration. *IEEE Trans. Multimed.* 19, 5 (2017), 1100–1113.

Weiqing Min, Shuqiang Jiang, Shuhui Wang, Jitao Sang, and Shuhuan Mei. 2017b. A delicious recipe analysis framework for exploring multi-modal recipes with various attributes. In *Proceedings of the ACM on Multimedia Conference.* 402–410.

Weiqing Min, Shuqiang Jiang, Shuhui Wang, Ruihan Xu, Yushan Cao, Luis Herranz, and Zhiqiang He. 2017c. A survey on context-aware mobile visual recognition. *Multimed. Syst.* 23, 6 (2017), 647–665.

Zhao Yan Ming, Jingjing Chen, Yu Cao, Ciaran Forde, Chong Wah Ngo, and Tat Seng Chua. 2018. Food photo recognition for dietary tracking; System and experiment. In *Proceedings of the International Conference on Multi Media Modelling.* 129–141.

Mingui Sun, John D. Fernstrom, Wenyan Jia, Steven A. Hackworth, Ning Yao, Yuecheng Li, Chengliu Li, Madelyn H. Fernstrom, and Robert J. Sclabassi. 2010. A wearable electronic system for objective dietary assessment. *J. Amer. Diet. Assoc.* 110, 1 (2010), 45.

Mark Mirtchouk, Christopher Merck, and Samantha Kleinberg. 2016. Automated estimation of food type and amount consumed from body-worn audio and motion sensors. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing.* 451–462.

Ruiko Miyano, Yuko Uematsu, and Hideo Saito. 2012. Food region detection using bag-of-features representation and color feature. In *Proceedings of the International Conference on Computer Vision Theory and Applications.* 709–713.

Ole G. Mouritsen, Rachel Edwards-Stuart, Yong-Yeol Ahn, and Sebastian E. Ahnert. 2017. Data-driven methods for the study of food perception, preparation, consumption, and culture. *Frontiers in ICT* 4 (2017), 15.

Nitish Nag, Vaibhav Pandey, and Ramesh Jain. 2017a. Health multimedia: Lifestyle recommendations based on diverse observations. In *Proceedings of the ACM International Conference on Multimedia Retrieval.* 99–106.

Nitish Nag, Vaibhav Pandey, and Ramesh Jain. 2017b. Live personalized nutrition recommendation engine. In *Proceedings of the International Workshopon Multimedia for Personal Health and Health Care.* 61–68.

Medic Nenad, Ziauddeen Hisham, Suzanna E. Forwood, Kirsty M. Davies, Amy L. Ahern, Susan A. Jebb, Theresa M. Marteau, and Paul C. Fletcher. 2016. The presence of real food usurps hypothetical health value judgment in overweight people. *Eneuro* 3, 2 (2016).

Nestle, Rosenberg, Bieber, Rogers, Haas, Dwyer, and Sigman-Grant. 1998. Behavioral and social influences on food choice—Discussion. *Nutrit. Rev.* 56, 5 (1998), S64–S74.

M. Ng, T. Fleming, M. Robinson, B. Thomson, N. Graetz, C. Margono, E. C. Mullany, S. Biryukov, C. Abbafati, and S. F. Abera. 2014. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: A systematic analysis for the global burden of disease study 2013. *Lancet* 384, 9945 (2014), 766.

Duc Thanh Nguyen, Zhimin Zong, Philip O. Ogunbona, Yasmine Probst, and Wanqing Li. 2014. Food image classification using local appearance and global structural information. *Neurocomputing* 140 (2014), 242–251.

Nag Nitish, Pandey Vaibhav, Sharma Abhisaar, Lam Jonathan, Wang Runyi, and Jain Ramesh. 2017. Pocket dietitian: Automated healthy dish recommendations by location. In *Proceedings of the International Conference on Image Analysis and Processing.* 444–452.

Karin Nordstrom, Christian Coff, Hakan Jönsson, Lennart Nordenfelt, and Ulf Gorman. 2013. Food and health: Individual, cultural, or scientific matters? *Genes Nutrit.* 8, 4 (2013), 357.

Ferda Ofli, Yusuf Aytar, Ingmar Weber, Raggi Al Hammouri, and Antonio Torralba. 2017. Is Saki #Delicious?: The food perception gap on Instagram and its relation to health. In *Proceedings of the International Conference on World Wide Web.* 509–518.

Koichi Okamoto and Keiji Yanai. 2016. An automatic calorie estimation system of food images on a smartphone. In *Proceedings of the International Workshop on Multimedia Assisted Dietary Management.* 63–70.

Luciano Oliveira, Victor Costa, Gustavo Neves, Talmai Oliveira, Eduardo Jorge, and Miguel Lizarraga. 2014. A mobile, lightweight, poll-based food identification system. *Pattern Recog.* 47, 5 (2014), 1941–1952.

Kaoru Ota, Minh Son Dao, Vasileios Mezaris, and Francesco G. B. De Natale. 2017. Deep learning for mobile multimedia: A survey. *ACM Trans. Multimed. Comput., Commun., Applic.* 13, 3s (2017), 34:1–34:22.

Jaderick P. Pabico, Alona V. De Grano, and Alan L. Zarsuela. 2015. Neural network classifiers for natural food products. Retrieved from: *arXiv preprint arXiv:1507.02346* (2015).

Lili Pan, Samira Pouyanfar, Hao Chen, Jiaohua Qin, Shu Ching Chen, Lili Pan, Samira Pouyanfar, Hao Chen, Jiaohua Qin, and Shu Ching Chen. 2017. DeepFood: Automatic multi-class classification of food materials using deep learning. In *Proceedings of the IEEE International Conference on Collaboration and Internet Computing.* 181–189.

Paritosh Pandey, Akella Deepthi, Bappaditya Mandal, and N. B. Puhan. 2017. FoodNet: Recognizing foods using ensemble of deep networks. *IEEE Sig. Proc. Lett.* 24, 12 (2017), 1758–1762.

D. Pauly. 1986. A simple method for estimating the food consumption of fish populations from growth data and food conversion experiments. *Fish. Bull.* 84 (1986), 827–839.

Sri Vijay Bharat Peddi, Pallavi Kuhad, Abdulsalam Yassine, Parisa Pouladzadeh, Shervin Shirmohammadi, and Ali Asghar Nazari Shirehjini. 2017. An intelligent cloud-based data processing broker for mobile e-health multimedia applications. *Fut. Gen. Comput. Syst.* 66 (2017), 71–86.

Maiyaporn Phanich, Phathrajarin Pholkul, and Suphakant Phimoltares. 2010. Food recommendation system using clustering analysis for diabetic patients. In *Proceedings of the International Conference on Information Science and Applications.* 1–8.

Parisa Pouladzadeh, Pallavi Kuhad, Sri Vijay Bharat Peddi, Abdulsalam Yassine, and Shervin Shirmohammadi. 2016a. Food calorie measurement using deep learning neural network. In *Proceedings of the Instrumentation and Measurement Technology Conference.* 1–6.

Parisa Pouladzadeh and Shervin Shirmohammadi. 2017. Mobile multi-food recognition using deep learning. *ACM Trans. Multimed. Comput. Commun. Applic.* 13, 3s (2017), 1–21.

Parisa Pouladzadeh, Shervin Shirmohammadi, and Rana Al-Maghrabi. 2014. Measuring calorie and nutrition from food image. *IEEE Trans. Instrument. Measur.* 63, 8 (2014), 1947–1956.

Parisa Pouladzadeh, Shervin Shirmohammadi, and Abdulsalam Yassine. 2016b. You are what you eat: So measure what you eat! *IEEE Instrument. Measur. Mag.* 19, 1 (2016), 9–15.

Parisa Pouladzadeh, Abdulsalam Yassine, and Shervin Shirmohammadi. 2015. FooDD: An image-based food detection dataset for calorie measurement. In *Proceedings of the International Conference on Multimedia Assisted Dietary Management.*

M. Puri, Zhiwei Zhu, Qian Yu, and A. Divakaran. 2009. Recognition and volume estimation of food intake using a mobile device. In *Proceedings of the Workshop on Applications of Computer Vision.* 1–8.

Francesco Ragusa, Valeria Tomaselli, Antonino Furnari, Sebastiano Battiato, and Giovanni M. Farinella. 2016. Food vs non-food classification. In *Proceedings of the International Workshop on Multimedia Assisted Dietary Management.* 77–81.

Daniele Ravl, Benny Lo, and Guang Zhong Yang. 2015. Real-time food intake classification and energy expenditure estimation on a mobile device. In *Proceedings of the IEEE International Conference on Wearable and Implantable Body Sensor Networks.* 1–6.

Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You only look once: Unified, real-time object detection. In *Proceedings of the Conference on Computer Vision and Pattern Recognition.* 779–788.

Jaclyn Rich, Hamed Haddadi, and Timothy M. Hospedales. 2016. Towards bottom-up analysis of social food. In *Proceedings of the International Conference on Digital Health Conference.* ACM, 111–120.

M. Rohrbach, S. Amin, M. Andriluka, and B. Schiele. 2012. A database for fine grained activity detection of cooking activities. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 1194–1201.

M. Rosenbaum, M. Sy, K. Pavlovich, R. L. Leibel, and J. Hirsch. 2008a. Leptin reverses weight loss-induced changes in regional neural activity responses to visual food stimuli. *J. Clinic. Investig.* 118, 7 (2008), 2583–2591.

Adam Sadilek, Henry Kautz, Lauren Diprete, Brian Labus, Eric Portman, Jack Teitel, and Vincent Silenzio. 2017. Deploying nEmesis: Preventing foodborne illness by data mining social media. *AI Magazine* 38, 1 (2017), 37–48.

Sina Sajadmanesh, Sina Jafarzadeh, Seyed Ali Ossia, Hamid R. Rabiee, Hamed Haddadi, Yelena Mejova, Mirco Musolesi, Emiliano De Cristofaro, and Gianluca Stringhini. 2017. Kissing Cuisines: Exploring worldwide culinary habits on the web. In *Proceedings of the International World Wide Web Conference.* ACM, 1013–1021.

Amaia Salvador, Nicholas Hynes, Yusuf Aytar, Javier Marin, Ferda Ofli, Ingmar Weber, and Antonio Torralba. 2017. Learning cross-modal embeddings for cooking recipes and food images. In *Proceedings of the Conference on Computer Vision and Pattern Recognition.* 3020–3028.

Satoshi Sanjo and Marie Katsurai. 2017. Recipe popularity prediction with deep visual-semantic fusion. In *Proceedings of the ACM Conference on Information and Knowledge Management.* 2279–2282.

Giovanni Schiboni and Oliver Amft. 2018. Automatic dietary monitoring using wearable accessories. In *Seamless Healthcare Monitoring: Advancements in Wearable, Attachable, and Invisible Devices.* Springer International Publishing, 369–412.

Mike Schuster and Kuldip K. Paliwal. 1997. *Bidirectional Recurrent Neural Networks.* IEEE Press. 2673–2681.

Yagcioglu Semih, Erdem Aykut, Erdem Erkut, and Ikizler-Cinbis Nazli. 2018. RecipeQA: A challenge dataset for multimodal comprehension of cooking recipes. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing.*

J. Senthilnath, Akanksha Dokania, Manasa Kandukuri, K. N. Ramesh, Gautham Anand, and S. N. Omkar. 2016. Detection of tomatoes using spectral-spatial methods in remotely sensed RGB images captured by UAV. *Biosyst. Eng.* 146 (2016), 16–32.

Zagoruyko Sergey and Komodakis Nikos. 2016. Wide residual networks. In *Proceedings of the British Machine Vision Conference.*

Paul W. Sherman and Jennifer Billing. 1999. Darwinian gastronomy: Why we use spices. *Bioscience* 49, 6 (1999), 453–463.

Wataru Shimoda and Keiji Yanai. 2015. CNN-based food image segmentation without pixel-wise annotation. In *Proceedings of the International Conference on Image Analysis and Processing.* 449–457.

Wataru Shimoda and Keiji Yanai. 2016. Foodness proposal for multiple food detection by training of single food images. In *Proceedings of the International Workshop on Multimedia Assisted Dietary Management.* 13–21.

Thiago H. Silva, Pedro O. S. de Melo, Jussara Almeida, Mirco Musolesi, and Antonio Loureiro. 2014. You are what you eat (and drink): Identifying cultural boundaries by analyzing food & drink habits in foursquare. In *Proceedings of the International Conference on Weblogs and Social Media.*

Ashutosh Singla, Lin Yuan, and Touradj Ebrahimi. 2016. Food/non-food image classification and food categorization using pre-trained GoogLeNet model. In *Proceedings of the International Workshop on Multimedia Assisted Dietary Management.* 3–11.

Lillian Sonnenberg, Emily Gelsomin, Douglas E. Levy, Jason Riis, Susan Barraclough, and Anne N. Thorndike. 2013. A traffic light food labeling intervention increases consumer awareness of health and healthy choices at the point-of-purchase. *Prevent. Med.* 57, 4 (2013), 253–257.

L. B. Sorensen, P. Moller, A. Flint, M. Martens, and A. Raben. 2003. Effect of sensory perception of foods on appetite and food intake: A review of studies on humans. *Int. J. Obesity* 27, 10 (2003), 1152.

Charles Spence, Carmel A. Levitan, Maya U. Shankar, and Massimiliano Zampini. 2010. Does food color influence taste and flavor perception in humans? *Chemosens. Percept.* 3, 1 (2010), 68–84.

Nitish Srivastava and Ruslan Salakhutdinov. 2012. Multimodal learning with deep Boltzmann machines. In *Proceedings of the International Conference on Neural Information Processing Systems.* 2222–2230.

Sebastian Stein and Stephen J. McKenna. 2013. Combining embedded accelerometers with computer vision for recognizing food preparation activities. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing.* 729–738.

Eliza Strickland. 2018. 3 sensors to track every bite and gulp. *IEEE Spectrum* 55, 7 (2018), 9–10.

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 1–9.

Ryosuke Tanno, Koichi Okamoto, and Keiji Yanai. 2016. DeepFoodCam: A DCNN-based real-time mobile food recognition system. In *Proceedings of the International Workshop on Multimedia Assisted Dietary Management.* 89–89.

Chun-Yuen Teng, Yu-Ru Lin, and Lada A. Adamic. 2012. Recipe recommendation using ingredient networks. In *Proceedings of the ACM Web Science Conference.* 298–307.

Chakkrit Termritthikun, Paisarn Muneesawang, and Surachet Kanprachar. 2017. NU-InNet: Thai food image recognition using convolutional neural networks on smartphone. *J. Telecommun., Electron. Comput. Eng.* 9, 2–6 (2017), 63–67.

Trung Phan Thanh and Daniel Gatica-Perez. 2017. # Healthy# Fondue# Dinner: Analysis and inference of food and drink consumption patterns on Instagram. In *Proceedings of the International Conference on Mobile and Ubiquitous Multimedia.* 327–338.

Frances E. Thompson and Amy F. Subar. 2017. Dietary Assessment Methodology. In *Nutrition in the Prevention and Treatment of Disease* (4th Ed.). Academic Press, 5–48.

Christoph Trattner and David Elsweiler. 2017a. Food recommender systems: Important contributions, challenges and future research directions. Retrieved from: *arXiv preprint arXiv:1711.02760* (2017).

Christoph Trattner and David Elsweiler. 2017b. Investigating the healthiness of internet-sourced recipes: Implications for meal planning and recommender systems. In *Proceedings of the International World Wide Web Conference.* 489–498.

Christoph Trattner, Markus Rokicki, and Eelco Herder. 2017. On the relations between cooking interests, hobbies and nutritional values of online recipes: Implications for health-aware recipe recommender systems. In *Proceedings of the ACM Conference on User Modeling, Adaptation and Personalization.*

Akihiro Tsubakida, Sosuke Amano, Kiyoharu Aizawa, and Makoto Ogawa. 2017. Prediction of individual eating habits using short-term food recording. In *Proceedings of the Workshop on Multimedia for Cooking and Eating Activities in conjunction with the International Joint Conference on Artificial Intelligence.* 45–48.

J. V. Verhagen and L. Engelen. 2006. The neurocognitive bases of human multimodal food perception: Sensory integration. *Neurosci. Biobehav. Rev.* 30, 5 (2006), 613–50.

Tri Vu, Feng Lin, Nabil Alshurafa, and Wenyao Xu. 2017. Wearable food intake monitoring technologies: A comprehensive review. *Computers* 6, 1 (2017), 4.

Claudia Wagner, Philipp Singer, and Markus Strohmaier. 2014. The nature and evolution of online food preferences. *EPJ Data Sci.* 3, 1 (2014), 38.

Kayo Waki, Kiyoharu Aizawa, Shigeko Kato, Hideo Fujita, Hanae Lee, Haruka Kobayashi, Makoto Ogawa, Keisuke Mouri, Takashi Kadowaki, and Kazuhiko Ohe. 2015. DialBetics with a multimedia food recording tool, FoodLog: Smartphone-based self-management for type 2 diabetes. *J. Diab. Sci. Technol.* 9, 3 (2015), 534–540.

Huayang Wang, Weiqing Min, Xiangyang Li, and Shuqiang Jiang. 2016. Where and what to eat: Simultaneous restaurant and dish recognition from food image. In *Proceedings of the Pacific Rim Conference on Multimedia.* 520–528.

Liping Wang, Qing Li, Na Li, Guozhu Dong, and Yu Yang. 2008. Substructure similarity measurement in Chinese recipes. In *Proceedings of the ACM International Conference on World Wide Web.* 979–988.

Xin Wang, Devinder Kumar, Nicolas Thome, Matthieu Cord, and Frederic Precioso. 2015. Recipe recognition with large multimodal food dataset. In *Proceedings of the IEEE International Conference on Multimedia and Expo Workshops.* 1–6.

D. A. Williamson, H. R. Allen, P. D. Martin, A. J. Alfonso, B. Gerald, and A. Hunt. 2003. Comparison of digital photography to weighed and visual estimation of portion sizes. *J. Amer. Diet. Assoc.* 103, 9 (2003), 1139–1145.

Hui Wu, Michele Merler, Rosario Uceda-Sosa, and John R. Smith. 2016. Learning to make better mistakes: Semantics-aware visual food recognition. In *Proceedings of the ACM on Multimedia Conference.* 172–176.

Wen Wu and Jie Yang. 2009. Fast food recognition from videos of eating for calorie estimation. In *Proceedings of the IEEE International Conference on Multimedia and Expo.* 1210–1213.

Rong Xiang, Huanyu Jiang, and Yibin Ying. 2014. Recognition of clustered tomatoes based on binocular stereo vision. *Comput. Electron. Agricult.* 106 (2014), 75–90.

Haoran Xie, Lijuan Yu, and Qing Li. 2011. A hybrid semantic item model for recipe search by example. In *Proceedings of the IEEE International Symposium on Multimedia.* 254–259.

Ruihan Xu, Luis Herranz, Shuqiang Jiang, Shuang Wang, Xinhang Song, and Ramesh Jain. 2015. Geolocalized modeling for dish recognition. *IEEE Trans. Multimed.* 17, 8 (2015), 1187–1199.

Keiji Yanai and Yoshiyuki Kawano. 2015. Food image recognition using deep convolutional network with pre-training and fine-tuning. In *Proceedings of the IEEE International Conference on Multimedia and Expo Workshops.* 1–6.

Keiji Yanai, Ryosuke Tanno, and Koichi Okamoto. 2016. Efficient mobile implementation of a CNN-based object recognition system. In *Proceedings of the 24th ACM International Conference on Multimedia.* 362–366.

Gao Yang, Zhang Ning, Honghao Wang, Ding Xiang, Ye Xu, Guanling Chen, and Cao Yu. 2016. iHear food: Eating detection using commodity Bluetooth headsets. In *Proceedings of the 1st IEEE International Conference on Connected Health: Applications.*

Longqi Yang, Cheng Kang Hsieh, Hongjian Yang, John P. Pollak, Nicola Dell, Serge Belongie, Curtis Cole, and Deborah Estrin. 2017. Yum-Me: A personalized nutrient-based meal recommender system. *ACM Trans. Inform. Syst.* 36, 1 (2017), 7.

Shulin Yang, Mei Chen, Dean Pomerleau, and Rahul Sukthankar. 2010. Food recognition using statistics of pairwise local features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 2249–2256.

Ning Yu, Desislava Zhekova, Can Liu, and Sandra Kubler. 2013. Do good recipes need butter? Predicting user ratings of online recipes. In *Proceedings of the International Workshop on Cooking with Computers.*

Jiawei Zhang, Limeng Cui, Philip S. Yu, and Yuanhua Lv. 2017a. BL-ECD: Broad learning based enterprise community detection via hierarchical structure fusion. In *Proceedings of the ACM Conference on Information and Knowledge Management.* 859–868.

Mabel Mengzi Zhang. 2011. *Identifying the Cuisine of a Plate of Food.* Technical Report. University of California, San Diego.

Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun. 2017b. ShuffleNet: An extremely efficient convolutional neural network for mobile devices. Retrieved from: *CoRR* abs/1707.01083 (2017).

Xi-Jin Zhang, Yi-Fan Lu, and Song-Hai Zhang. 2016. Multi-task learning for food identification and analysis with deep convolutional neural networks. *J. Comput. Sci. Technol.* 31, 3 (2016), 489–500.

Jiannan Zheng, Z. Jane Wang, and Chunsheng Zhu. 2017. Food image recognition via superpixel based low-level and mid-level distance coding for smart home applications. *Sustainability* 9, 5 (2017), 856.

Feng Zhou and Yuanqing Lin. 2016. Fine-grained image classification by exploring bipartite-graph labels. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 1124–1133.
Fengqing Zhu, Marc Bosch, Insoo Woo, Sung Ye Kim, Carol J. Boushey, David S. Ebert, and Edward J. Delp. 2010. The use of mobile devices in aiding dietary assessment and evaluation. *IEEE J. Select. Topics Sig. Proc.* 4, 4 (2010), 756.
Yaohui Zhu and Shuqiang Jiang. 2018. Deep structured learning for visual relationship detection. In *the Association for the Advance of Artificial Intelligence.* 7623–7630.
Yuke Zhu, Ce Zhang, Christopher Ré, and Li Fei-Fei. 2015. Building a large-scale multimodal knowledge base for visual question answering. Retrieved from: *CoRR* abs/1507.05670 (2015).
Yu Xiao Zhu, Junming Huang, Zi Ke Zhang, Qian Ming Zhang, Tao Zhou, and Yong Yeol Ahn. 2013. Geography and similarity of regional cuisines in China. *Plos One* 8, 11 (2013), e79161.
Zhimin Zong, Duc Thanh Nguyen, Philip Ogunbona, and Wanqing Li. 2010. On the combination of local texture and global structure for food classification. In *Proceedings of the IEEE International Symposium on Multimedia.* 204–211.

Received September 2018; revised March 2019; accepted April 2019