Analyzing and Recognizing Food in Constrained and Unconstrained Environments



Paolo Napoletano paolo.napoletano@unimib.it University of Milano-Bicocca Milan, Italy Gianluigi Ciocca gianluigi.ciocca@unimib.it University of Milano-Bicocca Milan, Italy

Raimondo Schettini raimondo.schettini@unimib.it University of Milano-Bicocca Milan, Italy



Figure 1: Schema of our food recognition and analysis system.

ABSTRACT

Recently, Computer Vision based image analysis techniques have attracted a lot of attention because they are used to develop automatic dietary monitoring applications. Food recognition is a quite challenging task: it is a non-rigid object, and is characterized by intrinsic high iter- and intra-class variability. The proper design of a food recognition system based on Computer Vision should contain several analysis stages. This paper reports on the most recent solutions in the field of automatic food recognition using computer vision developed at the Imaging and Vision Laboratory in the last 12 years. We present and discuss the main solutions developed and results achieved for food localization, segmentation, recognition and analysis. Food localization and segmentation aim at identifying the regions in the image corresponding to food items, food recognition aims at labeling each food region with the identity of the

AIxFood '21, October 20, 2021, Virtual Event, China

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 978-1-4503-8673-9/21/10...\$15.00 https://doi.org/10.1145/3475725.3483624 depicted food, and food analysis aims at determining properties of the food such as its quantity or ingredients.

CCS CONCEPTS

Computing methodologies → Machine learning; Computer vision; Computer vision problems; • General and reference → Surveys and overviews.

KEYWORDS

Computer vision; artificial intelligence; machine learning; convolutional neural networks; semantic segmentation; food recognition; grocery product recognition.

ACM Reference Format:

Marco Buzzelli, Gianluigi Ciocca, Paolo Napoletano, and Raimondo Schettini. 2021. Analyzing and Recognizing Food in Constrained and Unconstrained Environments. In *Proceedings of the 3rd Workshop on AIxFood (AIxFood '21), October 20, 2021, Virtual Event, China.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3475725.3483624

1 INTRODUCTION

Health care on food and good practices in dietary behavior are drawing people's attention recently. Nowadays technology can support the users in keep tracks of their food consumption, and to increase the awareness in their daily diet by monitoring their

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

food habits. Conventionally, such procedure has been accomplished by exploring logs manually recorded each day, howbeit manual recording is error-prone due to several factors like the inability to estimate the food type and quantity, difficulty to provide the continuity of such a demanding task, and delayed reporting. On the other hand, computer vision algorithms have come as a remedy for a more accurate and user-friendly procedure by automatizing dietary monitoring. For example, one may simply take a picture of a plate of food using a smartphone, and the whole process towards measuring the total calorie in the plate can be achieved by a visual understanding framework [35, 39]. To be useful for dietary monitoring, food recognition systems should also be able to operate in "wild" environments such as restaurants, canteens, and such.

In recent years, many research works have demonstrated that machine learning and computer vision techniques can help to build robust and effective dietary monitoring systems by providing algorithms to recognize diverse foods automatically, estimate the food quantity, calories, and composition using diverse machine learning techniques, from traditional to deep learning ones (e.g. [1, 7, 23, 24, 27, 33, 34, 38, 41–43]).

Automatic food recognition and analysis algorithms are often exploited in mobile applications. Examples of such systems are FoodLog [28], DietCam [30], Menu-Match [6], and FoodCam [26]. The use of mobile applications, allow the users to take advantage of these algorithms anywhere (i.e. in-the-wild) in an easy and friendly way. Differently from the in-the-wild scenario, there are other scenarios where these recognition techniques can be extremely useful. For example in the case of the analysis of the food intake for hospitalized patients. In this scenario, it is very important to correctly estimate the quantity also exploiting dedicated devices and hardware [31]. Similarly, ad-hoc solutions can be designed for canteen scenario where the constraints of the specific location (limited menu items, regulated access and serving) can be leveraged to build robust recognition and analysis algorithms.

Notwithstanding the extensive applicability and usefulness of automatic food recognition and analysis, these tasks are quite challenging. This is mainly due to the intrinsic properties of the food items. For instance, food is a non-rigid object. It is characterized by intrinsic high intra-class variability where the same food can have a very diverse visual appearance in different images due to different preparations, placements in the plate, or acquisition point of view.

The design of a food recognition system employing visual analysis algorithms is quite complex. Several stages are required and each of them need to be carefully designed to overcome the many challenges that the food appearance and its acquisition condition pose. Figure 1 shows an example of a food recognition and analysis system that we have designed and that incorporates the relevant modules necessary for building, for example, a complete dietary monitoring system [18]. The system is designed to analyze images of food dishes taken with a generic mobile device in an uncontrolled environment (e.g. restaurant, home, outdoor, ...) or in a controlled one (e.g. a canteen, an hospital, ...) to retrieve information about the food being consumed. For example, it can be used to deliver information about the ingredients in the recognized food to be sure that no food allergens are present. It can return the recipe of the food for future replica at home, or it can be used to provide an estimate of the food quantity for calories quantification. The recognition system is composed of several processing modules that are common to other food recognition systems, and the global architecture can be considered as a blueprint for many recognition applications.

In this paper we want to illustrate our achievements in the field of food recognition and analysis. As the Imaging and Vision Laboratory (IVL)¹ at the University of Milano-Bicocca, we published more than 20 papers related to food as the results of more than 12 years of investigation in the field where we analyzed almost every aspect of the food recognition and analysis. We will use the schema is Figure 1 to outline our works and achievements at the different processing steps. We will highlights our relevant findings, insights, and the design decision that we made to tackle the different aspects of the design of a comprehensive system.

Specifically, the overall food recognition system involves three main processing steps: food localization or segmentation aimed at identifying the regions in the image corresponding to food items, food recognition that labels each food region with the identity of the depicted food, and food analysis that determines properties of the food. The system's design is built upon our previous experiences in food segmentation, classification, recognition, and analysis. It is mostly based on machine learning techniques that have been demonstrated very useful in different computer vision tasks such as object detection, pattern recognition and image understanding. Most of our solutions are based on Deep Convolutional Neural Networks (DCNNs or CNNs in brief) specifically designed for foodrelated tasks.

Each processing step is carried out by a dedicated software module. Food localization or segmentation could be omitted if recognition can be efficiently performed on the whole image. The overall system is modular in the sense that the processing steps can be adapted, optimized, and scaled depending on the specific application objective. This also allows us to perform the processing on the acquisition device (i.e. smartphone or tablet) and/or on a separate server in the cloud. We also have designed algorithms for grocery and consumer product recognition, that can be easily integrated in the overall system. Finally, for the food analysis step, we have examples of approaches for recognize the food state, estimate the amount of leftovers in a canteen scenario, and for analyzing and presenting dish recipes.

2 FOOD LOCALIZATION AND SEGMENTATION

Food segmentation is quite challenging since food is characterized by an intrinsic high intra-class variability, and low inter-class variability in food appearance. The same food can have a very different visual appearance due to the preparation, presentation and scene, while different food may have extremely similar visual characteristics because of similar ingredients.

In [15] we tackled the problem of food segmentation in image of trays taken in a canteen. Working within a specific scenario, allowed us to design an ad-hoc segmentation strategy of the plates on the trays and the food on them. Our segmentation pipeline exploited

¹www.ivl.disco.unimib.it

domain dependent constraints and was based on traditional, handcrafted, image analysis processing and algorithms. First the plates on the trays are located, then an unsupervised, color and texture based segmentation algorithm [4, 19] is used to over segment the tray. The plate locations and the over-segmentation are merged and geometric constraints are posed to locate the food on the plates. Our strategy worked well in this specific application scenario. For an unconstrained scenario, a more general approach is required. We then tackled the segmentation approach using deep learning techniques [5]. Specifically we tested different implementation of the DeepLab neural network [9] for semantic segmentation or food localization. The promising results, inspired us to exploit CNNs for food localization and leave the specific food recognition to another processing step in the system.

We also found out that food segmentation algorithms are susceptible to the problem of the different illumination conditions under with the images are taken [10]. This problem arises when the food recognition system operates in uncontrolled scenarios. Even though CNN-based segmentation algorithms generalize well on new scenario, if they are trained on images with standard illumination, the overall performance degrades when the illuminations are too different. For this reason we created a new dataset of food images with different illuminant applied to them on order to train more robust food segmentation and localization algorithms [3]. In the same work we also considered acquisition artifacts such as noise, blur, and compression quality as possible defects that should be considered. On this image dataset, we were able to achieve an accuracy of 0.813 for the semantic segmentation on the illuminantvarying images using our GUN network architecture [32] (0.890 on the pristine images), and an accuracy of 0.940 for food localization using the same architecture (0.953 on the pristine images). These results outperforms the other. We made available the dataset to the research community to be used as a benchmark for segmentation algorithms in the hope that it could be useful to design more robust solutions².

3 FOOD RECOGNITION

With respect to the food recognition task, we have investigated three main scenario: the recognition of cooked (or prepared) dishes, the recognition of grocery products, and the recognition of food states.

3.1 Prepared dishes

In [14] and [15] we addressed the problem of food recognition in a canteen scenario. The food is put on trays and users select their meal within a set of daily dishes. The standard cutlery, the trays and the limited variability of food items make it possible to design an effective recognition system using hand-crafted and learned-based features and methodologies. The food regions are detected using traditional segmentation algorithms, while food recognition is performed locally the detected regions using a set of hand-crafted and CNN-based features coupled either with a k-NN or SVM classifier. The use of classic classification strategies and features was due to the limited amount of images collected in the canteen. We achieved about 79% of food and tray recognition accuracy using CNN-based features. For the canteen scenario, we also created a new image dataset UNIMIB2016³. The dataset was created by collecting 1,027 canteen trays for a total of 3,616 food instances belonging to 73 food classes.

Although we achieve good results in the canteen scenario, we wanted to address the food recognition problem in a less constrained scenario. To do this we needed a large dataset of food items acquired in-the-wild. In [17] we introduced a very large and heterogeneous food database for food recognition and retrieval, obtained by carefully merging databases from the state-of-the-art and thus creating one of the largest food database available in the literature with 475 food classes and 247,636 images. This database, denoted as Food-475, is an evolution of the Food-524 database that we presented in [16]. Food-524 contains 524 food classes obtained by syntactically merging food class names of four existing databases. Food-475, instead, contains food classes obtained after applying a semi-automatic merging procedure that considers semantically equivalent food classes.

With the new dataset, we evaluated different CNN-based features learned from a CNN trained on different, publicly available, databases. Features learned by deep Convolutional Neural Networks have been recognized to be more robust and expressive than hand-crafted ones. Given a CNN architecture and a training procedure, the efficacy of the learned features depends on the domain-representativeness of the training examples. We defined the food-domain representativeness of different food databases in terms of the total number of images, number of classes of the domain and number of examples for class. Different features are then extracted from a CNN based on the ResNet-50 [21], and trained on food databases with diverse food-domain representativeness. We evaluated these features for the tasks of food classification and retrieval. Results demonstrated that features learned from CNNs are very robust and powerful, and that the features extracted from the Food-475 database exhibits large improvements in accuracy for food classification and recognition tasks. For recognition, we achieved a Top-1 accuracy of 81.59%, and a Top-5 accuracy of 95.50%. This confirms the need for larger food databases in order to tackle the challenges in food recognition. For this reason we made our food datasets publicly available: Food-524⁴, and Food-475⁵.

3.2 Grocery products

Object detection/recognition in grocery scenarios is very complex because object appearance is highly variable due to relevant changes in scale, pose, viewpoint, lighting conditions and occlusion. Another challenge is due to the large number of product categories and subcategories. This means that recognition strategies need to cope with the problem of coarse to fine grained classification of the product categories [20, 29].

In [8] we proposed an automated fine-grained classification strategy specifically designed for edible plants, fruits, and mushrooms. The recognition of such items can be exploited in smart kitchen appliances. For example, a smart refrigerator can monitor the availability of some fruits and vegetables in a diet. As a reference dataset,

²http://www.ivl.disco.unimib.it/activities/benchmarking-food-segmentation/

³http://www.ivl.disco.unimib.it/activities/food-recognition/

⁴http://www.ivl.disco.unimib.it/activities/food524db/

⁵http://www.ivl.disco.unimib.it/activities/food475db/

the VegFru [22] is used for experimentation. The classification strategy is built on CNNs trained to recognize the coarse labels present in the dataset.

In our last works [12, 13], we designed a multi-task learning framework to exploit the intrinsic coarse- to fine- grained nature of the grocery products. By leveraging the hierarchical product classification provided by the Grocery Store Dataset [29], we designed a CNN network to be used as feature extractor in a classification and retrieval paradigms. Our experiments showed that our network, although it exhibited very good performance on the dataset used for training, has a limited generalization capabilities to other product datasets with very different contents. Specifically, the features learned on the Grocery Store Dataset are less adapt in discriminating the images in the Freiburg Groceries Dataset [25]. This indicate a bias that can be attuned by considering an enriched product dataset composed of a larger set of different product images.

3.3 Food state recognition

Being able to fully describe a food and its states will enable the implementation of intelligent dietary monitoring systems supporting users in controlling their food intake. To this end, in [11], we introduced a new dataset containing 20 different food categories taken from fruits and vegetables at 11 different states ranging from solid, sliced to creamy paste. We experimented with most common Convolutional Neural Network architectures on three different recognition tasks: food categories, food states, and both food categories and states. We exploited deep features extracted from CNNs combined with Support Vector Machines (SVMs) as an alternative to the End-to-End classification. We also compared deep features with several hand-crafted features.

The experiments confirmed the robustness of deep features. They were able to outperform hand-crafted features on all the three classification tasks and whatever is the food category or food state considered. Moreover, we found that there is no significant advantages in combining hand-crafted features with learned ones. From our experiments, emerged that the best features were those extracted from an Inception-v3 network: an accuracy of 92.89% for the food recognition task across states; 95.28% for the state recognition task across food categories; 90.71% for the combined food and state recognition task. We also very good results on the features extracted from a MobileNet-v2 network. This network could be a suitable alternative for mobile applications.

4 FOOD ANALYSIS

The aims of food analysis is the extraction of additional information about the identified food. Examples of information are: food ingredients, nutrients, allergenes, calories, and quantity [36, 37, 40]. These information are essential for a proper monitoring of the food intakes, especially for people in need for a controlled and healthy diet. In the following sub-sections we describe our works in food analysis related to leftover estimation, and recipe analysis.

4.1 Leftover estimation

In [14] we presented a dataset used for testing a system that recognizes foods and estimates food leftovers. The dataset contains 2,000 images of 15 classes of foods placed on trays. The images were acquired in a real canteen location, and are paired with the corresponding leftover images acquired after the meals. The images are associated to a given canteen customer by using a QR code automatically generated by the dietary monitoring system on the customer's mobile. Thanks to the way it has been annotated, this dataset can be used for food segmentation, recognition and for leftover estimation. This dataset, is the first and only dataset containing both the annotation of food dishes and the corresponding leftovers.

The estimation is performed by first segmenting each food dishes, and then by applying a patch-based recognition approach. The list of the served food is used to match plates before and after meal limiting the search space of the leftover estimation. The estimation of the leftover quantity is performed by counting the food patches of each food. For each food, the ratio between the patches found after the mail against those found before the meal represents the amount of leftover for that food. In our experiments, our approach was capable of estimating the relative quantity of eaten food with an average error of about 15 percentage points.

4.2 Recipes analysis

We considered the problem of recipe analysis in the context of traditional Italian cuisine. In [2] we developed some recipe processing tools for the analysis and interpretation of recipes, to identify terms able to describe ingredients, actions, and order of actions, using tools specific to Natural Language Processing (NLP). We formalized a recipe structure in a document type definition (DTD) with which we are able to produce a structured document with all the recipe's procedures. This document can be used for the identification of patterns that traditional recipes followed, to retrieve similar recipes by method, number of steps or total time. We designed also a prototype visualization tool for recipes using an interactive, graph-based, interface that allow the users to navigate and follow, step-by-step, the recipe. The prototype visualization application allows the user to easily access every step of the recipe procedure while having an overall overview of its components.

5 CONCLUSIONS

In this paper we present and discuss the main results we achieved in the field of automatic food recognition using Computer Vision in the last 12 years. Food recognition is a quite challenging task because food is characterized by intrinsic high intra-class variability where the same food can have a very diverse visual appearance in different images due to different preparations, placements in the plate, or acquisition point of view. We present an ideal pipeline of a food recognition and analysis system that contains several processing modules. We show how a complete food recognition and analysis pipeline can be implemented exploiting our previous works. Despite the good results of our approaches, robust food recognition is still an open problem, and are studying new solutions. Another topic that we are currently investigating is related to the food volume/quantity estimation. From the volume estimation, we can compute the calories intake of the dishes and thus monitoring the user diet in a more precise manner. We are working on novel approaches leveraging depth and contextual information.

REFERENCES

- Marios M Anthimopoulos, Lauro Gianola, Luca Scarnato, Peter Diem, and Stavroula G Mougiakakou. 2014. A food recognition system for diabetic patients based on an optimized bag-of-features model. *IEEE J. of Biomedical and Health Informatics* 18, 4 (2014), 1261–1271.
- [2] Maria Teresa Artese, Gianluigi Ciocca, and Isabella Gagliardi. 2021. Analysis of Traditional Italian Food Recipes: Experiments and Results. In *Pattern Recognition*. *ICPR International Workshops and Challenges*. Springer International Publishing, Cham, 677–690.
- [3] Sinem Aslan, Gianluigi Ciocca, Davide Mazzini, and Raimondo Schettini. 2020. Benchmarking Algorithms for Food Localization and Semantic Segmentation. International Journal of Machine Learning and Cybernetics 11 (2020), 2827–2847.
- [4] Sinem Aslan, Gianluigi Ciocca, and Raimondo Schettini. 2017. On Comparing Color Spaces for Food Segmentation. In New Trends in Image Analysis and Processing – ICIAP 2017: ICIAP International Workshops, WBICV, SSPandBE, 3AS, RGBD, NIVAR, IWBAAS, and MADiMa 2017, Catania, Italy, September 11-15, 2017, Revised Selected Papers, Sebastiano Battiato, Giovanni Maria Farinella, Marco Leo, and Giovanni Gallo (Eds.). Springer International Publishing, Cham, 435–443.
- [5] Sinem Aslan, Gianluigi Ciocca, and Raimondo Schettini. 2018. Semantic food segmentation by deep convolutional neural networks. In International Conference on Consumer Electronics - Berlin (ICCE-Berlin). IEEE, 1–6.
- [6] Oscar Beijbom, Neel Joshi, Dan Morris, Scott Saponas, and Siddharth Khullar. 2015. Menu-Match: Restaurant-Specific Food Logging from Images. In Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on. IEEE, 844–851.
- [7] M. Bosch, Fengqing Zhu, N. Khanna, C.J. Boushey, and E.J. Delp. 2011. Combining global and local features for food identification in dietary assessment. In *Image Processing (ICIP), 2011 18th IEEE International Conference on.* 1789–1792.
- [8] Marco Buzzelli, Federico Belotti, and Raimondo Schettini. 2018. Recognition of Edible Vegetables and Fruits for Smart Home Appliances. In 2018 IEEE 8th International Conference on Consumer Electronics-Berlin (ICCE-Berlin). IEEE, 1–4.
- [9] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence* 40, 4 (2017), 834–848.
- [10] Gianluigi Ciocca, Davide Mazzini, and Raimondo Schettini. 2019. Evaluating CNN-Based Semantic Food Segmentation Across Illuminants. In Computational Color Imaging (Lecture Notes in Computer Science, Vol. 11418), Shoji Tominaga, Raimondo Schettini, Alain Trémeau, and Takahiko Horiuchi (Eds.). Springer International Publishing, 247–259.
- [11] Gianluigi Ciocca, Giovanni Micali, and Paolo Napoletano. 2020. State Recognition of Food Images using Deep Features. IEEE Access 8 (2020), 32003–32017.
- [12] Gianluigi Ciocca, Paolo Napoletano, and Simone Giuseppe Locatelli. 2021. Iconic-Based Retrieval of Grocery Images via Siamese Neural Network. In Pattern Recognition. ICPR International Workshops and Challenges. Springer International Publishing, Cham, 269–281.
- [13] Gianluigi Ciocca, Paolo Napoletano, and Simone Giuseppe Locatelli. 2021. Multitask Learning for Supervised and Unsupervised Classification of Grocery Images. In Pattern Recognition. ICPR International Workshops and Challenges. Springer International Publishing, Cham, 325–338.
- [14] Gianluigi Ciocca, Paolo Napoletano, and Raimondo Schettini. 2015. Food Recognition and Leftover Estimation for Daily Diet Monitoring. In New Trends in Image Analysis and Processing – ICIAP 2015 Workshops (Lecture Notes in Computer Science, Vol. 9281), Vittorio Murino, Enrico Puppo, Diego Sona, Marco Cristani, and Carlo Sansone (Eds.). Springer International Publishing, 334–341.
- [15] Gianluigi Ciocca, Paolo Napoletano, and Raimondo Schettini. 2017. Food recognition: a new dataset, experiments and results. *IEEE Journal of Biomedical and Health Informatics* 21, 3 (2017), 588–598.
- [16] Gianluigi Ciocca, Paolo Napoletano, and Raimondo Schettini. 2017. Learning CNN-based Features for Retrieval of Food Images. In New Trends in Image Analysis and Processing – ICIAP 2017: ICIAP International Workshops, WBICV, SSPandBE, 3AS, RGBD, NIVAR, IWBAAS, and MADiMa 2017, Catania, Italy, September 11-15, 2017, Revised Selected Papers, Sebastiano Battiato, Giovanni Maria Farinella, Marco Leo, and Giovanni Gallo (Eds.). Springer International Publishing, Cham, 426–434.
- [17] Gianluigi Ciocca, Paolo Napoletano, and Raimondo Schettini. 2018. CNN-based Features for Retrieval and Classification of Food Images. *Computer Vision and Image Understanding* 176–177 (2018), 70–77.
- [18] Gianluigi Ciocca, Paolo Napoletano, and Raimondo Schettini. 2018. IVLFood-WS: Recognizing food in the wild using Deep Learning. In International Conference on Consumer Electronics - Berlin (ICCE-Berlin). IEEE, 1–6.
- [19] Y. Deng and B.S. Manjunath. 2001. Unsupervised segmentation of color-texture regions in images and video. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, 8 (2001), 800–810.

- [20] Annalisa Franco, Davide Maltoni, and Serena Papi. 2017. Grocery product detection and recognition. Expert Systems with Applications 81 (2017), 163–176.
- [21] 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.
- [22] Saihui Hou, Yushan Feng, and Zilei Wang. 2017. Vegfru: A domain-specific dataset for fine-grained visual categorization. In *Proceedings of the IEEE International Conference on Computer Vision*. 541–549.
- [23] Takumi Ege and Keiji Yanai. 2018. Multi-task learning of dish detection and calorie estimation. In Proceedings of the Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management. ACM, 53–58.
- [24] Shuqiang Jiang, Weiqing Min, Linhu Liu, and Zhengdong Luo. 2019. Multi-Scale Multi-View Deep Feature Aggregation for Food Recognition. *IEEE Transactions* on Image Processing 29 (2019), 265–276.
- [25] Philipp Jund, Nichola Abdo, Andreas Eitel, and Wolfram Burgard. 2016. The freiburg groceries dataset. arXiv preprint arXiv:1611.05799 (2016).
- [26] Yoshiyuki Kawano and Keiji Yanai. 2014. FoodCam: A real-time food recognition system on a smartphone. *Multimedia Tools and Applications* (2014), 1–25.
- [27] Chairi Kiourt, George Pavlidis, and Stella Markantonatou. 2020. Deep learning approaches in food recognition. In *Machine Learning Paradigms*. Springer, 83– 108.
- [28] 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 2009 Workshop on Multimedia for Cooking and Eating Activities. 23–30.
- [29] Marcus Klasson, Cheng Zhang, and Hedvig Kjellström. 2019. A hierarchical grocery store image dataset with visual and semantic labels. In 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 491–500.
- [30] Fanyu Kong and Jindong Tan. 2012. DietCam: Automatic dietary assessment with mobile camera phones. *Pervasive and Mobile Computing* 8, 1 (2012), 147–163.
- [31] Ya Lu, Thomai Stathopoulou, Maria F Vasiloglou, Stergios Christodoulidis, Zeno Stanga, and Stavroula Mougiakakou. 2020. An artificial intelligence-based system to assess nutrient intake for hospitalised patients. *IEEE transactions on multimedia* 23 (2020), 1136–1147.
- [32] Davide Mazzini and Schettini Raimondo. 2019. Spatial Sampling Network for Fast Scene Understanding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 97–107.
- [33] Simon Mezgec and Barbara Koroušić Seljak. 2017. Nutrinet: A deep learning food and drink image recognition system for dietary assessment. *Nutrients* 9, 7 (2017), 657.
- [34] Weiqing Min, Shuqiang Jiang, Linhu Liu, Yong Rui, and Ramesh Jain. 2019. A survey on food computing. ACM Computing Surveys (CSUR) 52, 5 (2019), 1–36.
- [35] A. Myers, N. Johnston, V. Rathod, A. Korattikara, A. Gorban, N. Silberman, S. Guadarrama, G. Papandreou, J. Huang, and K. Murphy. 2015. Im2Calories: Towards an Automated Mobile Vision Food Diary. In *IEEE Int. Conf. on Computer Vision (ICCV)*. 1233–1241.
- [36] Robin Ruede, Verena Heusser, Lukas Frank, Alina Roitberg, Monica Haurilet, and Rainer Stiefelhagen. 2021. Multi-Task Learning for Calorie Prediction on a Novel Large-Scale Recipe Dataset Enriched with Nutritional Information. In 2020 25th International Conference on Pattern Recognition (ICPR). IEEE, 4001–4008.
- [37] 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 IEEE Conference on Computer Vision and Pattern Recognition. 3068–3076.
- [38] M. A. Subhi, S. H. Ali, and M. A. Mohammed. 2019. Vision-Based Approaches for Automatic Food Recognition and Dietary Assessment: A Survey. *IEEE Access* 7 (2019), 35370–35381.
- [39] Ryosuke Tanno, Koichi Okamoto, and Keiji Yanai. 2016. Deepfoodcam: A dcnnbased real-time mobile food recognition system. In Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management. ACM, 89–89.
- [40] Quin Thames, Arjun Karpur, Wade Norris, Fangting Xia, Liviu Panait, Tobias Weyand, and Jack Sim. 2021. Nutrition5k: Towards Automatic Nutritional Understanding of Generic Food. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 8903–8911.
- [41] Wen Wu and Jie Yang. 2009. Fast food recognition from videos of eating for calorie estimation. In *Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on.* IEEE, 1210–1213.
- [42] Shulin Yang, Mei Chen, Dean Pomerleau, and Rahul Sukthankar. 2010. Food recognition using statistics of pairwise local features. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2249–2256.
- [43] Ning Yao, Robert J Sclabassi, Qiang Liu, Jie Yang, John D Fernstrom, Madelyn H Fernstrom, and Mingui Sun. 2007. A video processing approach to the study of obesity. In *Multimedia and Expo, 2007 IEEE International Conference on*. IEEE, 1727–1730.