Ontology-Driven Food Category Classification in Images

Ivan Donadello[1][0000−0002−0701−5729] and Mauro Dragoni[1][0000−0003−0380−6571]

Fondazione Bruno Kessler, Via Sommarive 18, I-38123, Trento, Italy
{donadello,dragoni}@fbk.eu

Abstract. The self-management of chronic diseases related to dietary habits includes the necessity of tracking what people eat. Most of the approaches proposed in the literature classify food pictures by labels describing the whole recipe. The main drawback of this kind of strategy is that a wrong prediction of the recipe leads to a wrong prediction of any ingredient of such a recipe. In this paper we present a multi-label food classification approach, exploiting deep neural networks, where each food picture is classified with labels describing the food categories of the ingredients in each recipe. The aim of our approach is to support the detection of food categories in order to detect which one might be dangerous for a user affected by chronic disease. Our approach relies on background knowledge where recipes, food categories, and their relatedness with chronic diseases are modeled within a state-of-the-art ontology. Experiments conducted on a new publicly released dataset demonstrated the effectiveness of the proposed approach with respect to state-of-the-art classification strategies.

Keywords: Food Classification · Knowledge-Based System · Food Tracking · Food Dataset · mHealth.

1 Introduction

Chronic diseases are responsible for approximately 70% of deaths among Europe and U.S. each year and they account for about 75% of the health spending[1],[2]. Such chronic diseases can be largely preventable by eating healthily, exercising regularly, avoiding (tobacco) smoking, and receiving preventive services. Prevention at every stage of life would help people stay healthy, avoid or delay the onset of diseases, and keep diseases they already have from becoming worse or debilitating; it would also help people lead productive lives and, at the end, reduce the costs of public health.

Dietary tracking is one of the pillars for the self-management of chronic diseases. One of the most common modalities for tracking eaten food is to keep a diary of food pictures as implemented in numerous commercial applications. The use of food pictures opens the challenge of recognizing all the taken food from

---

[1] http://www.who.int/nmh/publications/ncd_report_full_en.pdf
[2] https://www.cdc.gov/media/releases/2014/p0501-preventable-deaths.html

users' pictures. State-of-the-art approaches classify meal images according to the food they contain. However, they are not able to infer the food categories given by the recipe of that particular food. The detection of these categories is fundamental for people affected by particular diseases, such as diabetes, hypertension or obesity.

In this work, we propose a strategy based on the multi-label classification of food pictures according to the food categories contained in a specific food recipe of the Mediterranean diet. We compare this method against the (more standard) single-label classification of the food recipe and the inference of the contained food categories. Our claim is that a classification error in a single food recipe affects the majority of the inferred food categories. For example, a single-label classifier may confuse two similar pasta recipes, e.g., "Pasta with carbonara sauce" and "Pasta with cheeses" that, even if they might be aesthetically similar, they have different food categories. Indeed, the former contains cold cuts that affect people suffering of cardiovascular diseases. This can be prevented with a multi-label classification. Moreover, food categories, thanks to the use of background knowledge, can be associated with a risk level with respect to specific diseases. Within this scenario, the use of background knowledge helps for two reasons. Firstly, it gives the possibility of modeling logical relationships between food categories and risk levels with respect to specific diseases. Secondly, information collected from users can be exploited within a behavior change context to support them in changing their dietary habits through the implementation of goal-based strategies [21]. The contribution of the paper is the following:

- food pictures are classified with respect to the set of food categories contained in the food recipe. This outperforms the standard (single-label) classification of food recipes and the consequent inference of the food categories;
- background knowledge is used for inferring which are the food categories contained within each recipe together with information about the risk level of each food category with respect to a first identified set of chronic diseases;
- a new dataset of food pictures and the source code of the classification tool have been released in order to support the reproducibility of the results and to foster further research in this direction.

2   Related Work

The recognition of foods from images is the first step for the dietary tracking. This task has been studied by the Computer Vision community with techniques of image classification/segmentation and volume estimation. The first works rely on the extraction of visual features from the images and the consequent use of classifiers. The main features used are local and global features, SIFT, textons and local binary patterns [9,11,13,12,15,17,1]. The classifiers are k-NN classifiers, Support Vector or Kernel Machines. The works in [13,12] also developed the first food images datasets: the Food50 and Food85, with 50 and 85 labels of Japanese foods, respectively. In [15] the authors developed the UEC FOOD-100

dataset (100 food labels), successively extended with 256 labels in UEC FOOD-256 [17]. Food-101 [1] is one of the biggest datasets having 101,000 images with 101 food labels. Here the authors mine discriminant food image parts with Random Forests and classify them with a SVM. These techniques have been used in mobile apps for food tracking, such as, Food-Log [18], DietCam [19], Food-Cam [17], Snap-n-Eat [29]. They also perform an approximate estimation of the taken calories with volume estimation techniques. However, the growing availability of huge datasets and hardware resources has made Convolutional Neural Networks (CNNs) the standard technique for food classification [4,22,28,14,16,2], thus avoiding the use of engineered features. In [2] the authors combine CNNs and Conditional Random Fields to predict both food and ingredient labels in a multi-task learning setting. They also developed one of the biggest food images dataset: the VIREO Food-172 dataset. It contains 172 food labels, 353 ingredients labels and 110,241 images. The Food524DB dataset is used in [4] for food recognition with CNNs and gathers the Food50, Food-101, UEC FOOD-256 and VIREO Food-172 datasets. It contains 524 food labels and 247,636 images.

A more fine-grained analysis of the meal is performed by estimating also the quantity of food in the dish and the consequent calories intake. The first step is the semantic segmentation of the food in the dish and the quantity computation with techniques of volume estimation. However, these techniques also require a database of foods and relatives densities. The GoCARB [5] system estimates the carbohydrates intake for people with diabetes. After a segmentation of the foods their classification is performed with SVMs. The volume estimation performs the 3D reconstruction of the food with stereo vision techniques. A density table returns the carbohydrates for each food label. A similar technique is used also in [26]. Other works exploit a known reference object (e.g., a thumb [25] or a wallet [24]) for volume computation or assume a defined shape template for a given class of foods [10]. The Im2Calories system [23] uses a CNN to predict a depth map of the food image that is used to build the 3D model of the meal. Quantity estimation can be addressed with a multi-task learning approach by defining a tailored CNN that both learns the classification of the food in the dish and the relative calories or volume. However, this interesting direction requires a dataset with the annotated calories [8] or the depth information in the images [20]. In [3] the authors use CNNs to perform semantic segmentation to estimate the leftovers in the trays in the canteens. They also developed the UNIMIB2016 dataset with 73 food labels to test their method.

Few works among those mentioned above predict food categories and match them with some nutritional facts in a database [8,25,29,5,2]. Also, they predict only one food category (e.g., pasta) for each detected food and this can be inaccurate. Indeed, a pasta dish should be avoided by a person suffering of diabetes. However, a pasta dish might have carbonara sauce, containing eggs, aged cheese and cold cuts. One, or more, of these food categories could be not suitable for people suffering of obesity, hypertension or cardiovascular diseases. In these cases it is important to have a food recognition system that performs multi-label classification of the several food categories in the dish.

3   Background Knowledge

The use of background knowledge allows the design of intelligent systems having the purpose of going beyond the sole classification of food images. Such background knowledge, indeed, enables the possibility of exploiting logic relationships and inference capabilities for reusing the results of the food classification task in order to support users for more complex goals. For example, background knowledge can formalize specific dietary patterns that can be used to improve users' lifestyle, avoiding the rise or sharpening of chronic diseases, and to support them in changing their behaviors. Here we propose a strategy to predict food categories from food images. These categories might represent a warning for people affected by specific diseases (e.g. "Pasta" for people affected by diabetes). Our approach relies on a state-of-the-art conceptual model for the Mediterranean diet, called HeLiS, defining the dietary and physical activity domains together with entities modeling concepts concerning users' profiles and the monitoring of their activities. For the description of the conceptual model and of the methodology adopted for building it, the reader can refer to [7]. Here, the HeLiS ontology (`http://w3id.org/helis`) has been extended by adding, to the dietary domain, information concerning the risk level of food categories with respect to specific diseases. In this section, we limit to mention the main concepts involved into the food classification task proposed in this paper together with the ones modeled within the HeLiS ontology extension. Figure 1 shows an excerpt of the HeLiS ontology containing the concepts involved into our classification task.
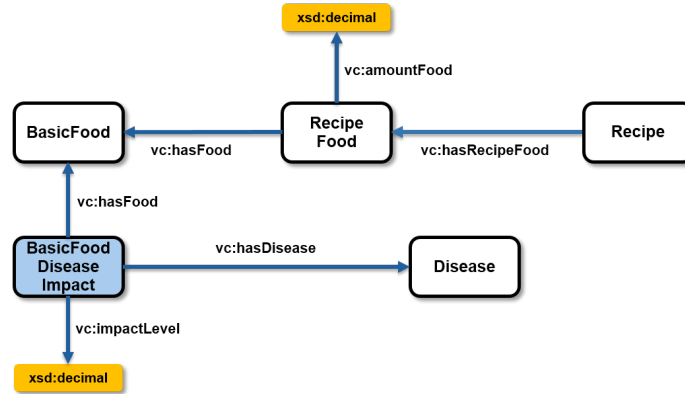

Fig. 1. Excerpt of the HeLiS ontology including the main concepts exploited for the proposed food classification approach.

Instances of the `BasicFood` concept describe foods for which micro-information concerning nutrients (carbohydrates, lipids, proteins, minerals, and vitamins) is available. Moreover, these instances also contain information about the category to which each `BasicFood` belongs to (such as `Pasta`, `Aged Cheese`, `Eggs`, `Cold`

`Cuts` and `Vegetal Oils`). While instances of the `Recipe` concept describe the composition of complex dishes (such as `Pasta with Carbonara Sauce`) as a list of instances of the `RecipeFood` concepts. This concept reifies the relationships between each `Recipe` individual, the list of `BasicFood` it contains and the amount of each `BasicFood`. Besides this dual classification, instances of both `BasicFood` and `Recipe` concepts are categorized under a more fine-grained structure. Regarding the number of individuals, currently, the `HeLiS` ontology contains 986 individuals of type `BasicFood` and 4408 individuals of type `Recipe`.

The `Disease` concept models the chronic diseases supported by the system and for which information about the risk level relationship with specific `BasicFood` is available. Currently, we instantiated the `Disease` concept for the "diabetes", "kidney diseases", "cardiovascular diseases", "hypertension", and "obesity" diseases. Finally, the `BasicFoodDiseaseImpact` concept reifies the relationships between each `Disease` and `BasicFood` individuals and, for each reification, it contains a number representing the risk level of that `BasicFood` for that `Disease`. The risk level is represented by a numeric value ranging from 0 (no risk) to 3 (high risk) and is useful for the generation of warning messages to users in a behaviuoral-change system. For example, the food category `Eggs` has a low risk level for diabetes. Thus, the warning messages for a user suffering of diabetes will be soft if the user exceeds with the consume of eggs.

4   Multi-label Food Category Classification

Our goal is to assign every food image with a set of food category labels. These categories refer to the ingredients that compose the food recipe in the image and are provided by the `HeLiS` ontology. We address this problem as a multi-label image classification task where $\mathcal{X} \in \mathbb{R}^d$ is the input domain of our images and `BasicFood` is the set of the possible food category labels. Given an image $\boldsymbol{x} \in \mathcal{X}$, we need to predict a vector $\boldsymbol{y} = \{y_1, y_2, \ldots, y_K\} \subseteq$ `BasicFood` where y_i is the i-th food category label associated to $\boldsymbol{x}$. Up to our knowledge, state-of-the-art methods in food image recognition do not exploit multi-label classification. They classify images according to only one single label taken from `Recipe`. Therefore, we exploit two methods: i) a direct multi-label classification of the food categories with a CNN and ii) a single-label image classification of the food recipes (e.g., `Pasta with Carbonara Sauce`) with a CNN and then the logical inference of all its food categories (i.e., `Pasta`, `Eggs`, etc) through the `RecipeFood` concept.

4.1   Methods

Current methods in image classification use supervised deep learning techniques based on CNNs. These are able to learn the salient features of an image in order to classify it according to some training examples. Many CNNs have been developed exploiting several combinations of the hidden layers (convolutions, poolings, activations) in order to address the main challenges of the visual recognition. In both methods i) and ii) we separately train (on the dataset in Section 5.1) one

of the most performing CNN, the Inception-V3 [27]. This network presents convolutional filters of multiple sizes operating at the same level. This makes the network "wider" and able to better detect the salient parts of an image both locally and globally. Finally, the network has a standard fully-connected layer for predicting the classes.

Direct Multi-Label Classification For this task we train the Inception-V3 for directly learning the vector $\boldsymbol{y}$ of the food categories in `BasicFood`. We use a sigmoid as activation function of the last fully-connected layer and binary cross entropy as loss function. This is a standard setting for multi-label classification.

Single-Label Classification and Inference Another method to classify the food categories in a meal image consists in: i) classifying an input image with a CNN according to the food label it contains (e.g., `Pasta with Carbonara Sauce`). This is the standard multiclass classification where one image is classified with only one food label among many classes. ii) Inferring the food category labels from the food label by using the concepts and properties of `HeLiS`. In our example, the detection of `Pasta with Carbonara Sauce` implies the presence of these food categories: `Pasta`, `Eggs`, `Aged cheese`, `Vegetal Oils` and `Cold cuts`. More formally, let CNN an Inception-V3 trained to multiclassify food labels in `Recipe`. Here the activation function of the last fully-connected layer is a softmax and the loss function is a categorical cross entropy. Thus $CNN(\boldsymbol{x}) = \langle s_1, s_2, \ldots, s_n \rangle$ with $s_i \in \mathbb{R}$ is the classification score of the network for the label $l_i \in$ `Recipe`. Let $l^* \in$ `Recipe` be the label with highest score in $CNN(\boldsymbol{x})$, then the food category labels vector $\boldsymbol{y}$ is defined as:

$$\boldsymbol{y} = \{y_i \in \texttt{BasicFood} \mid \exists w \in \texttt{RecipeFood} : \texttt{hasFood}(w, y_i) \wedge$$
$$\texttt{hasRecipeFood}(l^*, w)\} \quad (1)$$

5 Experiments

Here we compare the multi-label and single-label plus inference methods for the food category classification from meal images. Our claim is that a classification error in a single food recipe affects the majority of the inferred food categories leading to inaccurate results. The dataset and the tool are publicly available at `https://github.com/ivanDonadello/Food-Categories-Classification`.

5.1 The Food and Food Categories (FFoCat) Dataset

The `HeLiS` ontology contains the food and food category concepts (Section 3) exploited in the multi-label classification. We build a new dataset from these concepts. We sample some of the most common recipes in `Recipe` and use them as food labels. The food categories are then automatically retrieved from `BasicFood` with a SPARQL query. Examples of food labels are `Pasta with Carbonara`

`Sauce` and `Baked Sea Bream`. Their associated food categories are `Pasta, Aged Cheese, Vegetal Oils, Eggs, Cold Cuts` and `Fresh Fish, Vegetal Oils`, respectively. We collect 156 labels for foods and 51 for food categories. We scrape the Web using Google Images as search engine to download all the images related to the food labels. Then, we manually clean the dataset resulting in 58,962 images with 47,108 images for the training set and 11,854 images for the test set (80-20 ratio of splitting). The dataset is affected by some natural imbalance, indeed the food categories present a long-tail distribution: only few food categories labels have the majority of the examples. On the contrary, many food categories labels have few examples. This makes the food classification challenging.

5.2   Experimental Settings

For both multi and single-label we train the Inception-V3 network from scratch on the FFoCat training set (with different loss functions) to find the best set of weights. The fine tuning using a pre-trained Inception-V3 did not perform sufficiently. We resized the images to 299x299 pixels and perform data augmentation by using rotations, width and height shifts, shearing, zooming and horizontal flipping. We run 100 epochs of training with a batch size of 16 and a learning rate of 10^{-6}. We adopt the early stopping criterion to prevent overfitting. The training has been performed with the Keras framework (TensorFlow as backend) on a PC equipped with a NVIDIA GeForce GTX 1080 Ti. We obtain the 93.43% and the 41.02% of accuracy for the multi and single-label classification tasks.

5.3   Metrics

As performance metric we use the mean average precision (MAP) that summarizes the classifier precision-recall curve. This is computed by listing the obtained classification scores of the food/food categories for all the test set pictures. We threshold this list at multiple values in $[0, 1]$ and the predictions are the set of labels with score higher than the threshold. The MAP is $\sum_{i=1}^{n}(R_n - R_{n-1})P_n$, i.e., the weighted mean of precision P_n achieved at each threshold level n. The weight is the increase of the recall in the previous threshold: $R_n - R_{n-1}$. The macro AP is the average of the AP over the classes, the micro instead considers each entry of the predictions as a label. We prefer MAP instead of accuracy as the latter can give misleading results for sparse vectors. Indeed, Accuracy = (TP+TN)/(TP+TN+FP+FN) with TP (TN) the true positives (negatives) and FP (FN) the false positives (negatives). Therefore, a classifier returning a zero vector y for the 51 food categories achieves an accuracy of 92%.

5.4   Results and Discussions

Given an (set of) input image(s) x, the computing of the precision-recall curve requires the predicted vector y of food category labels and a score associated to each label in y. In the multi-label method this score is directly returned by the

Inception-V3 network. In the single-label and inference method this score needs to be computed. We tested two strategies: i) we perform *exact inference* of the food categories from HeLiS and assign the value 1 to the scores of each $y_i \in \boldsymbol{y}$; ii) the food categories labels inherit the *uncertainty* returned by the CNN. The score of each y_i is the value s_i returned by $CNN(\boldsymbol{x})$. Table 1 reports the results.

Table 1. The multi-label classification of food categories outperforms in average precision (AP) the methods based on single-label classification and logical inference.

Method	Micro-AP (%)	Macro-AP (%)
Multi-label	**76.28**	**49.81**
Single-class exact	50.74	31.82
Single-class uncert.	60.51	42.73

The direct multi-label model outperforms the single-label models of approximately 26 and 16 points of micro-AP and 21 and 8 points of macro-AP, respectively. The micro-AP is always better than the macro-AP as it is sensible to the mentioned imbalance of the data. Moreover, the precision-recall curve (Figure 2) of the direct multi-label model is always above the other models. This confirm our claim that errors in the single recipe classification propagate to the majority of the food categories the recipe contains. That is, the inferred food categories will be wrong because the recipe classification is wrong. On the other hand, errors in the direct multi-label classification will affect only few food categories. Good per-
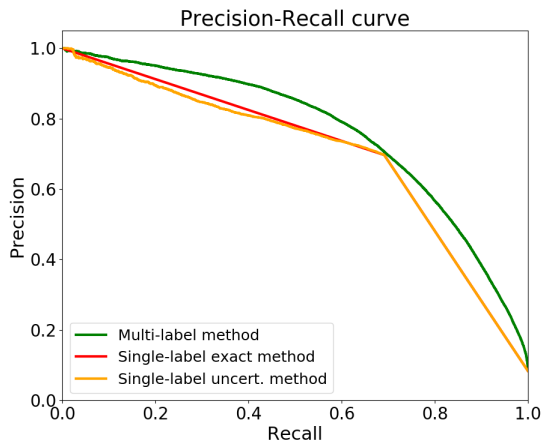

Fig. 2. The multi-label classification of food categories outperforms in average precision (AP) the methods based on single-label classification and logical inference.

formance in dietary-tracking systems are important especially if the predictions are used in a behavioural-change system for generating proper user feedback. Indeed, the misclassification of a meal could generate wrong warning messages or even no message to users. To this aim, we also perform a qualitative comparison of the methods using testing images, see Figure 3. The top-left meal of Figure 3

Fig. 3. Example images leading to wrong user messages with the multi-class model.
.

contains a `Pasta with Garlic, Oil and Chili Peppers` that is misclassified by the single-label method with a `Pasta with Carbonara Sauce`, thus inferring wrong `Eggs` and `Cold Cuts`. In this case, for example, the intake of `Cold Cuts` could violate a dietary restriction (e.g., to consume no more than two portions of cold cuts in a week) with the consequent generation of an erroneous warning message for a user that should avoid the excessive intake of `ColdCuts`. Here, the multi-label method classifies all the categories correctly. The top-right image contains a `Vegetable Pie`, the single-label method misclassifies it and infers the wrong category of `Pizza Bread`, whereas the multi-label method is more precise. The low-left image contains `Backed Potatoes` and the single-label classification classifies it as `Backed Pumpkin` thus missing the category of `Fresh Starchy Vegetables`. This category is retrieved by the multi-label method that, within a behavioural-change system, can trigger the generation of a warning message for people affected by, for example, diabetes. Regarding the last low-right image, the single-label classification and inference method wrongly classifies the input image as `Tomato and Ricotta Cheese Pasta`, thus containing `FreshCheese` instead of `Eggs` and `TomatoSauces` instead of `ColdCuts`. In this case, no warning message will be generated for a user that should avoid `ColdCuts` and has already violated the consequent restriction in the last few days.

6 Conclusions

This paper discusses a multi-label food classification strategy for classifying food pictures based on the food categories contained in the recipe instead of the recipe

itself. The aim of the proposed approach is to detect food categories having a high risk level for people affected by specific chronic diseases. The proposed strategy relies on the use of background knowledge exploited for inferring food categories from a recipe and their links with the risk level associated with each chronic disease. Moreover, we provide a new dataset containing 58,962 annotated images. Results demonstrated the effectiveness of the proposed classification strategy. Moreover, our proposal outperforms a more standard method based on single-image classification and inference of the food categories.

Future work will focus on designing multi-task learning algorithms for the joint prediction of both foods and food categories. In addition, we want to further exploit the combination of deep learning with ontologies by using constraints-based methods, such as Logic Tensor Networks [6], already applied to image classification tasks. Both these directions will be tested on bigger and standard image datasets containing food and food categories, such as VIREO FOOD-172 [2]. Finally, the proposed strategy opens also the possibility of being integrated into intelligent systems implementing behavior change policies for supporting users in adopting healthy lifestyles.

References

1. Bossard, L., Guillaumin, M., Gool, L.J.V.: Food-101 - mining discriminative components with random forests. In: ECCV (6). Lecture Notes in Computer Science, vol. 8694, pp. 446–461. Springer (2014)
2. Chen, J., Ngo, C.: Deep-based ingredient recognition for cooking recipe retrieval. In: ACM Multimedia. pp. 32–41. ACM (2016)
3. Ciocca, G., Napoletano, P., Schettini, R.: Food recognition: A new dataset, experiments, and results. IEEE J. Biomedical and Health Informatics **21**(3), 588–598 (2017)
4. Ciocca, G., Napoletano, P., Schettini, R.: Learning cnn-based features for retrieval of food images. In: ICIAP Workshops. Lecture Notes in Computer Science, vol. 10590, pp. 426–434. Springer (2017)
5. Dehais, J., Anthimopoulos, M., Shevchik, S., Mougiakakou, S.G.: Two-view 3d reconstruction for food volume estimation. IEEE Trans. Multimedia **19**(5), 1090–1099 (2017)
6. Donadello, I., Serafini, L., d'Avila Garcez, A.S.: Logic tensor networks for semantic image interpretation. In: IJCAI. pp. 1596–1602. ijcai.org (2017)
7. Dragoni, M., Bailoni, T., Maimone, R., Eccher, C.: Helis: An ontology for supporting healthy lifestyles. In: International Semantic Web Conference (2). Lecture Notes in Computer Science, vol. 11137, pp. 53–69. Springer (2018)
8. Ege, T., Yanai, K.: Image-based food calorie estimation using knowledge on food categories, ingredients and cooking directions. In: ACM Multimedia (Thematic Workshops). pp. 367–375. ACM (2017)
9. Farinella, G.M., Moltisanti, M., Battiato, S.: Classifying food images represented as bag of textons. In: ICIP. pp. 5212–5216. IEEE (2014)
10. He, Y., Xu, C., Khanna, N., Boushey, C.J., Delp, E.J.: Food image analysis: Segmentation, identification and weight estimation. In: ICME. pp. 1–6. IEEE Computer Society (2013)

11. He, Y., Xu, C., Khanna, N., Boushey, C.J., Delp, E.J.: Analysis of food images: Features and classification. In: Image Processing (ICIP), 2014 IEEE International Conference on. pp. 2744–2748. IEEE (2014)
12. Hoashi, H., Joutou, T., Yanai, K.: Image recognition of 85 food categories by feature fusion. In: ISM. pp. 296–301. IEEE Computer Society (2010)
13. Joutou, T., Yanai, K.: A food image recognition system with multiple kernel learning. In: ICIP. pp. 285–288. IEEE (2009)
14. Kagaya, H., Aizawa, K., Ogawa, M.: Food detection and recognition using convolutional neural network. In: ACM Multimedia. pp. 1085–1088. ACM (2014)
15. Kawano, Y., Yanai, K.: Real-time mobile food recognition system. In: CVPR Workshops. pp. 1–7. IEEE Computer Society (2013)
16. Kawano, Y., Yanai, K.: Food image recognition with deep convolutional features. In: UbiComp Adjunct. pp. 589–593. ACM (2014)
17. Kawano, Y., Yanai, K.: Foodcam-256: A large-scale real-time mobile food recognition system employing high-dimensional features and compression of classifier weights. In: ACM Multimedia. pp. 761–762. ACM (2014)
18. Kitamura, K., Yamasaki, T., Aizawa, K.: Foodlog: capture, analysis and retrieval of personal food images via web. In: CEA@ACM Multimedia. pp. 23–30. ACM (2009)
19. Kong, F., Tan, J.: Dietcam: Automatic dietary assessment with mobile camera phones. Pervasive and Mobile Computing 8(1), 147–163 (2012)
20. Lu, Y., Allegra, D., Anthimopoulos, M., Stanco, F., Farinella, G.M., Mougiakakou, S.G.: A multi-task learning approach for meal assessment. In: MADiMa@IJCAI. pp. 46–52. ACM (2018)
21. Maimone, R., Guerini, M., Dragoni, M., Bailoni, T., Eccher, C.: Perkapp: A general purpose persuasion architecture for healthy lifestyles. Journal of Biomedical Informatics 82, 70–87 (2018)
22. Mezgec, S., Koroušić Seljak, B.: Nutrinet: a deep learning food and drink image recognition system for dietary assessment. Nutrients 9(7), 657 (2017)
23. Myers, A., Johnston, N., Rathod, V., Korattikara, A., Gorban, A.N., Silberman, N., Guadarrama, S., Papandreou, G., Huang, J., Murphy, K.: Im2calories: Towards an automated mobile vision food diary. In: ICCV. pp. 1233–1241. IEEE Computer Society (2015)
24. Okamoto, K., Yanai, K.: An automatic calorie estimation system of food images on a smartphone. In: MADiMa @ ACM Multimedia. pp. 63–70. ACM (2016)
25. Pouladzadeh, P., Shirmohammadi, S., Almaghrabi, R.: Measuring calorie and nutrition from food image. IEEE Trans. Instrumentation and Measurement 63(8), 1947–1956 (2014)
26. Puri, M., Zhu, Z., Yu, Q., Divakaran, A., Sawhney, H.S.: Recognition and volume estimation of food intake using a mobile device. In: WACV. pp. 1–8. IEEE Computer Society (2009)
27. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: CVPR. pp. 2818–2826. IEEE Computer Society (2016)
28. Zhang, W., Zhao, D., Gong, W., Li, Z., Lu, Q., Yang, S.: Food image recognition with convolutional neural networks. In: UIC/ATC/ScalCom. pp. 690–693. IEEE Computer Society (2015)
29. Zhang, W., Yu, Q., Siddiquie, B., Divakaran, A., Sawhney, H.: snap-n-eat food recognition and nutrition estimation on a smartphone. Journal of diabetes science and technology 9(3), 525–533 (2015)