Food9K: Detecting and Localizing Food in Social Media Images Using YOLOv3

Mohammad Akbari, Zahra Golpayegani
Amirkabir University of Technology
{akbari.ma,zahragolpayegani}@aut.ac.ir

ABSTRACT

Social media platforms like Instagram, Twitter, and Pinterest provide a virtually unlimited source of data to study food consumption behavior in everyday life. In this paper, we present Food9K, a new large-scale dataset of food images crawled from social media. Food9K images are labeled with class names and bounding box coordinates that specify locations of food items present in each image. Using this dataset, we train a YOLO model to localize and detect food items in unseen data. This model can further be used to analyze food consumption behavior from social media data.

1 INTRODUCTION

Food is an essential part of our life, health, and well-being. Hence, understanding people's food intake and diet has been studied from both individual aspects, such as food choice [25] and food perception [32], and community aspects, such as food safety [5] and food culture [14]. With the advancement of mobile and wearable devices, several applications have been developed to record our daily food intake via a personal food log system [11]. These log systems analyze recorded data and make recommendations for healthier diet practice, such as calorie measurement [26], computer-aided diet assessment [20], and nutrition balance estimation [31].

Nowadays, social networking services (SNSs) have been integrated into our daily lifestyle, where millions of users share posts about what happened around them, and diet is no exception [16]. Instagram as a photo-sharing platform is more common as people can share multi-modal information in one post, i.e., photos or videos with geo-tag information or descriptive captions and hashtags. Further, users can receive comments and likes on their posts from their friend circle. As a vital part of our life, food also is a prevalent theme on Instagram; 69% of young adults take a picture or video of their food before eating, and food fans connect to Instagram 18 times a day[1] on average. While social media data is increasingly studied to learn users' interests and behaviors [10, 12], limited research investigates the available social data as an important data source for understanding users diet and food consumption patterns.

Despite its value and significance, user-generated content (UGC) on social networks has not been fully utilized due to the following challenges. First, social media data can be extremely noisy because it is mostly user-generated, and users can freely express themselves. Thus, filtering noisy data is a significant challenge when obtaining social media data, and accurate noise-filtering is crucial when working with social media data. Second, how to crawl relevant data from social media is another unavoidable challenge. To collect food-related posts, we need to create a food dictionary as a reference. We can then collect the posts that contain at least one of the food dictionary items in their caption. Therefore, we need to create a list of food items that can give us the most relevant data to use. Last but not least, we would like to mention an intrinsic characteristic of food images shared on social media. Such images are captured in-the-wild; there is not necessarily one food item in each image. The location of the food item within the image is not known beforehand, the lighting properties vary from image to image, and there may be some noises obstacles in the picture, etc. These properties harden image recognition and localization tasks while making this dataset of in-the-wild captured images more valuable for general studies.

In this paper, we exploit the availability of food images on social media to collect a large-scale food image dataset. Currently-available food datasets are small in either number of categories or number of images per class. Therefore, they have limited capability for building an accurate food recognition model. Furthermore, most of the food datasets available do not give us information about the location of the food item within the image. Food9K dataset addresses both of these issues by first, being a relatively large food dataset in the number of classes and the number of images per category. Second, localizing food items in an image using bounding boxes. Therefore, Food9K makes real-time food detection and localization possible for a wide variety of dishes.

To summarize, our contributions are as follows:

- We propose a framework for collecting food-related posts from social media using a food dictionary.
- We used our data collection system to collect Food9K, a dataset of social media images and their metadata such as caption, hashtags, number of likes, and comments.
- We trained a YOLO model on the UECFOOD-256 dataset and used it to semi-automate the labeling process for the new Food9K dataset.
- After labeling Food9K, we trained a YOLO model on Food9K that can perform food detection (recognition and localization) on new unseen data.

2 BACKGROUND AND RELATED WORK

We divide this section into four sub-sections to analyze related work in more depth:

2.1 Food Datasets

There are several food datasets currently available to the researchers, each of which has its own characteristics. One of the most important benchmark food datasets available online is the ETHZ Food-101 dataset, which is introduced in [3]. ETHZ Food-101 is a dataset of 101 food categories with 101,000 food images (1000 images per

[1] https://blog.digimind.com/en/trends/instagram-key-global-figures-2019

category). The food categories were the 101 most popular categories in a food website, and images were downloaded from the same website. This dataset was introduced to study automatic food recognition using Random Forests. Their model outperformed other classification methods except for CNN.

UECFOOD256 [19] dataset contains photos of 256 food categories. This dataset is valuable for finding the location of each food item in an input image because each food photo has a bounding box indicating the location of the food item presented in the photo. The main objective of creating this dataset was to use it for implementing a practical food recognition system that can detect common food types in Japan in photos taken with Smart Phones.

Some food datasets were created from existing ones, such as Food524DB [8], which is a dataset of 524 food categories constructed by merging four benchmark datasets: VIREO, Food-101, Food50, and a modified version of UECFOOD256. The authors used this dataset to fine-tune a Residual Network for the task of food image recognition. They also created Food475DB [9] dataset, which is obtained by semantically merging equivalent food classes of the Food524DB dataset, continuing the previous work. Food475DB is a dataset of 475 food classes with 247,636 images.

In [6], the authors introduced the UNIMIB2015 food dataset, a relatively small dataset with only 15 food categories, aimed to be used for food recognition and leftover estimation. This dataset is composed of full tray images and leftover tray images, with a total of 2000 images. In a later work, authors created the UNIMIB2016 [7] dataset, which is a more diverse version of UNIMIB2015, composed of 1,027 tray images with multiple food items and containing 73 food categories. This dataset is only used for food recognition and does not include leftover trays.

Some scholars have studied regional cuisine as well. In [4], Chen et al. introduced a large-scale food image dataset called ChineseFoodNet for automatic dish recognition using a deep convolutional neural network (CNN). This dataset contains 180,000 food photos of 208 categories, all Chinese food.

Cookpad Image Dataset [13] is a dataset of more than 1.64 million images after cooking, including more than 3.10 million pictures taken during the cooking process collected from Cookpad, a recipe search service. This dataset is one of the largest food datasets and includes images during the cooking process as well as after the cooking has finished. Also, each image in this dataset is linked to a corresponding recipe corpus. No food image recognition model was applied to Cookpad Image Dataset by the authors of this study.

In a similar work, Salvador et al. introduced Recipe1M+ [22], a large-scale dataset of over one million cooking recipes and 13 million food images collected from cooking websites. However, this dataset is designed for image-recipe retrieval and not for food image recognition.

The characteristics of the Food9K dataset that make it unique are that first, the data source is social media, and images are taken in-the-wild. Second, this dataset contains images from a wide variety of dishes, and each class has enough data to train a food classifier with acceptable precision. Finally, Food9K is labeled with bounding boxes that can help deep learning classifiers localize food items within each image.

2.2 Food Dictionaries

Depending on the platform, there are several ways to search for food-related posts on social media. On Instagram, for example, you can search for posts by location or hashtag. To find food-related posts, searching by hashtag is a common method, which requires a reference food-item list. Having prepared such a list, one can find food-related posts by simply searching for social media posts that contain at least one of the food list items in their hashtags. We call this food-item list **"Food Dictionary"**.

There are several ways to make a food dictionary. In [24], the food dictionary of 1302 items was created by manually analyzing the menus of restaurants in Abu Dhabi and adding some food-related words. This list is not limited to dishes and some food-related words such as "dining-out", "eat out", and "fast-food" are also included.

In [30], the food dictionary was obtained from an online food vocabulary word list, containing 564 food items. This list includes fruits and cooking utensils as well as some food-related words such as "boil", "fried", and "lunch box". Similar to this work, in [33], the authors utilized a basic food vocabulary list that is used for teaching English. This list contains general food words such as "meat", "fish", and "cake".

In [17], they created the food dictionary by asking Kenyan experts to provide them with a list of 38 foods that are popular in Kenya. This list is in the Kiswahili language. Authors used this list to search for Instagram posts that included at least one of the 38 words in their image captions.

In this work, we are looking for a food dictionary that only includes dish names, i. e., we do not want to have names of any fruits, desserts, beverages, or cooking utensils in our dictionary. Furthermore, we are looking for popular dishes that are recognizable by people of all countries. Thus, our dictionary will be written in English, and no dish name specific to one country is included. Also, to be able to use this dictionary in practice, we should use specific dish names, such as "meatballs", rather than listing general food-related words, e.g., "meat".

2.3 Visual Food Recognition

The task of identifying food items in an image has been a challenge for a while. Although some studies have focused on single-label food recognition, multi-label food recognition has been more popular in recent years. Among all the approaches to tackle this challenge, CNN-based approaches are proven to be more effective than conventional methods [18].

Different types of networks have been used for food recognition so far. In [15], authors evaluated the effectiveness of classifying food images using Google's image detection architecture, Inception. They fine-tuned this architecture for classifying food images in three datasets: ETH Food-101, UECFOOD-100, and UECFOOD-256. They achieved top-1 accuracy of 88.28%, 81.45%, and 76.17% on these datasets, respectively.

In another work [20], the authors proposed a new CNN-based approach inspired by AlexNet and GoogleNet architectures for food image recognition and applied it to UECFOOD256 and Food-101 food image datasets and achieved the top-1 accuracy of 76.3% and 77.4% on each dataset respectively.

In [21], they compared the performance of the bag-of-features (BoF) model coupled with a support vector machine (SVM) with a five-layer CNN on a small-scale dataset of 5822 food images over ten food categories. They achieved a much better accuracy using the five-layer CNN and improved the results using data augmentation techniques based on geometric transformations.

Food recognition on social media noisy data has also been studied. In [30], Sharma et al. used Instagram's official API to extract nutritional information from Instagram posts. In [17], authors trained a classifier to recognize 13 popular food types in Kenya on a dataset collected from Instagram. They used this dataset to study food trends in Kenya. In [1], authors collected food images from Twitter and used a CNN network to identify the categories of food that people share on social media. Using this data, they created World-FoodMap, a map that visualizes the popularity and trends of food all over the world.

To train a highly accurate food recognition model, we need to have a big-scale dataset of food images. Currently-available datasets fail to provide researchers enough data to train a practical food recognition model. However, thanks to the data boom on social media platforms such as Instagram, we can use food-related posts uploaded by users to create an extended food dataset that can recognize a wide variety of food items.

2.4 Visual Food Localization

Only a few papers have been published on food localization, and none of them have used social media as their data source. As long as we know, this paper addresses the problem of food localization on social media for the first time.

In [23], a mobile system is presented to localize the meal portion of a food image and report its nutritional information. However, the photos are from 23 specific restaurants, and the size of each portion is known beforehand.

In [2], the authors have proposed the first method for simultaneous food localization and recognition using a heat map of probabilities. This method can draw bounding boxes around each food item in an image. However, it is not focused on food images shared on social media.

In this study, we use YOLOv3 to train a model on the Food9K dataset that can detect and localize food items in an input image.

3 FOOD9K DATASET

To the best of our knowledge, there is no prior work on building a large-scale image dataset based on social media data, which provides necessary (meta-)data for data-driven research on social media and people's food consumption behavior. In this section, we introduce a data collecting framework to build Food9K. See Figure 1 for examples of images in Food9K with their labeled bounding boxes. As dataset labeling and noise filtering is a costly step in building large-scale dataset, we also describe how to accelerate these steps.

3.1 Data Collection

To download food-related posts from social media, we first need to have a representative list of daily food items, i.e., a food dictionary. To construct this list, we used the DOHMH MenuStat database [2], a public database of menu information from top American restaurant chains provided by the New York City Department of Health and Mental Hygiene. First, five categories of food were excluded from the database in order to consider dishes only. These categories are Beverages, Toppings & Ingredients, Appetizers and Sides, Desserts, and Baked Goods. Then, the resulting data was sorted by popularity in decreasing order, and the top 1000 food items were extracted from the database.

This list of 1000 most popular food items needed further processing to represent only dish names. Some words such as numbers (e.g., "eggs for 2"), descriptive adjectives (e.g., "homestyle turkey"), or restaurant names were omitted from food items for this study. This process resulted in having duplicate food items that were deleted in the next step. Next, we grouped food items of similar dishes and chose the most descriptive name. Consider "Pizza" as an example. Different kinds of pizzas were included in the list, such as "Mexican Pizza", "Veggie Pizza", "Sausage Pizza", "Meats Pizza", "Chicken Pizza", "Margherita Pizza", "Pepperoni Pizza", "Cheese Pizza", etc. We chose "Pizza" from this group as a representative name and removed other variations from the list to have a diverse list of food names.

After processing the initial list, we finally ended up with a food dictionary of 174 dish names. Using our food dictionary, We searched for social media data and downloaded photos that contained at least one of the food dictionary items in their captions. We also collected metadata for each downloaded image. This metadata consists of image captions, hashtags, timestamps, number of likes, and the number of food words in image captions that were also listed on our food dictionary. We collected up to 2000 images per category shared on social media between 2017/3/1 to 2020/3/1, which resulted in a total of 335899 images. However, the downloaded dataset suffered from noise and needed data cleaning.

3.2 Data Cleaning

Since the dataset was collected based on keyword searches, it contains lots of noisy data, i.e., non-food images with food words in their captions. Traditionally, manual labeling through crowdsourcing is used for data annotation. However, this method is not efficient for building a large scale dataset, e.g., social media data. Thus we adopted an alternative method to filter out noisy data. We leveraged a pre-trained classifier to detect non-food images as proposed in [27]. More specifically, we considered the output of the food/non-food classifier as the label of the images and removed all images tagged as non-food. To evaluate the accuracy of the final dataset selected by this process, we manually calculated the accuracy of a randomly selected subset of our dataset. Based on our calculation, the food/non-food model achieved a high accuracy of 96.86% on noisy social media data.

After discarding non-food images, we removed categories with less than 500 images from our dataset. We finally ended up with a dataset of 9287 images over 35 food categories, called **Food9K**.

[2] https://data.cityofnewyork.us/Health/DOHMH-MenuStat/qgc5-ecnb

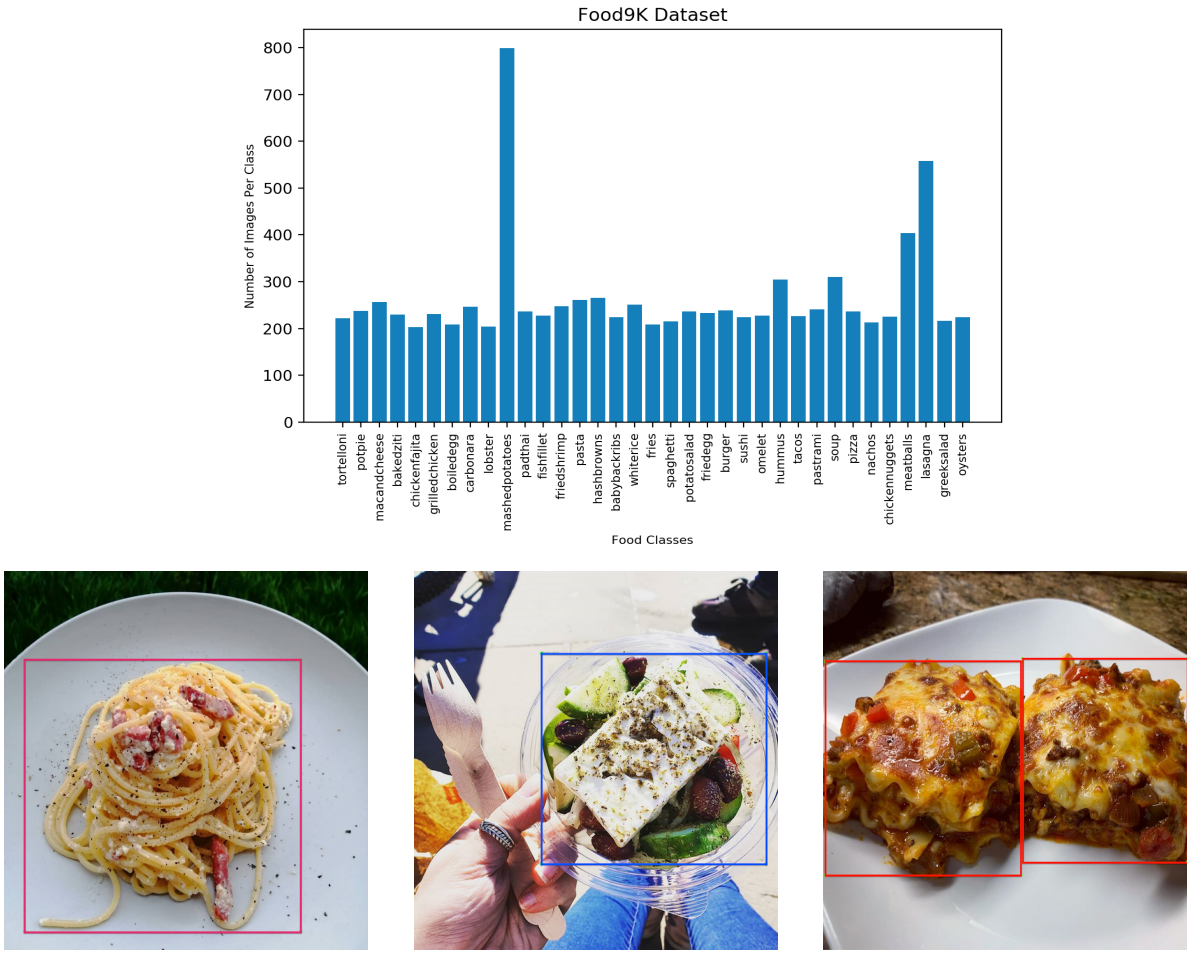

Figure 1: Examples of food images and their bounding boxes in Food9K dataset. left: Carbonara, middle: Greek Salad, right: Lasagnas

3.3 Data Labeling

Drawing bounding boxes and labeling class names for more than 9K images by hand is a tedious process. To accelerate labeling, we trained an object detection model on the UECFOOD-256 dataset. Then, we applied this model to our Food9K dataset to have a machine-generated bounding box for each food item in each image (see Figure 2 and 3). In the next step, we manually went through all the pictures and examined bounding boxes for each image one by one and added class names to each bounding box in the same run. Using this technique, we labeled the Food9K dataset in a semi-automated way.

4 METHODOLOGY

Having our dataset labeled with bounding box coordinates and class names, we can start training our model. To detect and localize food items in an image, we used a popular object detection model, YOLO.

4.1 Training the Model

We used YOLO (You Only Look Once) [28], a state-of-the-art, real-time object detection system to simultaneously perform food localization and detection. YOLO uses convolutional neural networks for object detection, and it is one of the fastest object detection algorithms available. YOLO, as a detection algorithm, detects locations of objects in the form of bounding boxes as well as class labels. YOLO divides the image into several grids and draws a fixed number of anchors for each grid to decide whether an object's center falls inside an anchor or not, measured by a probability between 0 to 1. It will then output those bounding boxes that have a probability higher than a threshold value. The image is fed to CNN only once; hence, this algorithm performs much faster than its competitors.

4.2 Experimental Setting

The dataset is randomly split into training and test sets with 7430 images in the training set (80.1% of the dataset), and 1857 images in the test set (19.9%). For a fair comparison, all experiments were conducted using PyTorch deep learning framework. Training has been performed using stochastic gradient descent (SGD) with momentum, where the initial learning rate is set to 0.001, momentum is set

0.9, and weight decay is set to 0.0005. Anchors are recalculated for our dataset to achieve more accuracy. Training images are resized to 416x416, and only mosaic data augmentation was applied. We have fine-tuned the model we trained on the UECFOOD-256 on our food dataset.

4.3 Results

In order to evaluate our trained model, we need to have metrics for both food detection and food localization. We will use F1-score to evaluate detection and GIoU (Generalized Intersection over Union) loss [29] to evaluate localization. GIoU loss is defined as:

$$GIoU = IoU - \frac{|C \setminus A \cup B|}{|C|}$$

$$GIoU\,Loss = 1 - GIoU$$

Our approach achieved 0.633 as an F1-score and minimized the GIoU loss down to 1.67. By comparing the results of the two models, one trained on the UECFOOD-256 dataset and the other trained on the Food9K dataset, we can observe that we achieved a higher F1-score for the latter. See Table 1 for other evaluation metrics.

Table 1: Results of Evaluation

Dataset Name	Precision	F1-Score	GIoU
UECFOOD-256	0.517	0.516	0.924
Food9K	0.591	0.633	1.67

4.4 Discussion

We have created a benchmark food dataset and achieved a reasonable accuracy for detecting and localizing food images using our YOLO model. This model is a tool for analyzing food consumption behavior and studying food trends all around the world. It can also be used for diet assessment and nutrition estimation to help people choose what to eat to stay healthy. Since this model can localize food items, it can be used for volume estimation by applying a few improvements.

However, since Food9K is a dataset of 35 food categories, any model trained on this dataset will fail to detect some dishes that are not listed in our dataset. The proposed framework for collecting, cleaning, and labeling social media data can be repeated for any other food category to extend Food9K. We recommend using semi-supervised learning to label food images in a more efficient way.

Food detection and localization remain a challenging task for some reasons. First, dishes come in various shapes, textures, sizes, and colors because of using different ingredients in each recipe. Second, some dishes have several names, for example, "tacos dorados", "rolled taco", or "flauta" are all other names for "taquito". This challenge should be handled carefully during labeling. Third, some visually similar dishes are entirely different, e.g., Spaghetti and Noodles. Distinguishing these dishes may be difficult even for humans if the recipe is not available. Providing too much data is not an answer to this problem because it will cause overfitting.

bayad bashe in:

Figure 2: Example of an accurate bounding box created by the model trained on UECFOOD-256 on an unseen image from the FOOD9K dataset.

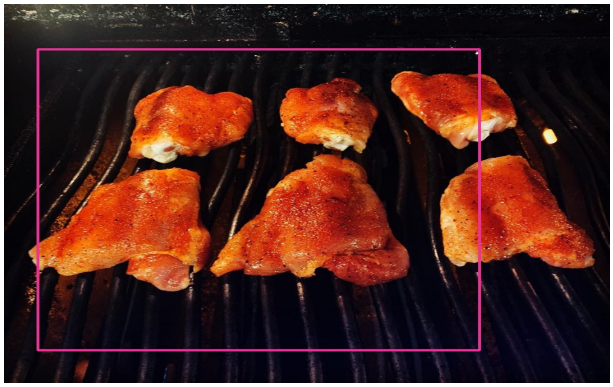

Figure 3: Example of an inaccurate bounding box created by the model trained on UECFOOD-256 on an unseen image from the FOOD9K dataset. Notice how the model failed to enclose all food items in the bounding box.

5 CONCLUSION

The importance of analyzing food intake led us to the problem of automatic food image detection and localization. On the other hand, people are used to sharing what happens around them on social media platforms, and food is not an exception. In this study, we proposed a framework to collect food data from social media, filter out noisy data, and label data using a semi-automated approach. We used this framework to create Food9K, a dataset of food images with their metadata collected from social media platforms. We have also trained a YOLOv3 model on this dataset that can detect and localize food items present in an input image with an F1-score of 0.633. To the best of our knowledge, this study performs food localization on social media data for the first time.

6 FUTURE WORK

Food9K dataset comes with some metadata for each photo that can be used for further research. The image caption, comments, number of likes, etc. can help the researchers answer a wide range of questions about food consumption behavior. Therefore, this dataset can be used for future research in many different areas. Some examples are as follows:

- Food recommendation application
- Diet planning application

- Investigating a possible relation between diabetes rate and calorie consumption in a specific city
- Studying a possible relation between weather conditions and calorie consumption
- Designing a caption recommender system based on a food image input

We are also planning to boost the accuracy of the model, experiment food detection, and localization using other architectures. Furthermore, we will include more dishes labeled with semi-supervised learning approaches in future works.

REFERENCES

[1] Giuseppe Amato, Paolo Bolettieri, Vinicius Monteiro de Lira, Cristina Ioana Muntean, Raffaele Perego, and Chiara Renso. 2017. Social Media Image Recognition for Food Trend Analysis. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR âĂŹ17)*. Association for Computing Machinery, New York, NY, USA, 1333âĂŞ1336. https://doi.org/10.1145/3077136.3084142

[2] Marc BolaÃśos and Petia Radeva. 2016. Simultaneous Food Localization and Recognition. (2016). arXiv:cs.CV/1604.07953

[3] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. 2014. Food-101–mining discriminative components with random forests. In *European conference on computer vision*. Springer, 446–461.

[4] Xin Chen, Hua Zhou, Yu Zhu, and Liang Diao. 2017. ChineseFoodNet: A large-scale Image Dataset for Chinese Food Recognition. *arXiv preprint arXiv:1705.02743* (2017).

[5] Zikuan Chen and Yang Tao. 2001. Food safety inspection using âĂIJfrom presence to classificationâĂİ object-detection model. *Pattern recognition* 34, 12 (2001), 2331–2338.

[6] 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)*, Vittorio Murino, Enrico Puppo, Diego Sona, Marco Cristani, and Carlo Sansone (Eds.), Vol. 9281. Springer International Publishing, 334–341. https://doi.org/10.1007/978-3-319-23222-5_41

[7] 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. https://doi.org/10.1109/JBHI.2016.2636441

[8] 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, 426–434. https://doi.org/10.1007/978-3-319-70742-6_41

[9] Gianluigi Ciocca, Paolo Napoletano, and Raimondo Schettini. 2018. CNN-based Features for Retrieval and Classification of Food Images. *Computer Vision and Image Understanding* - (2018), –. https://doi.org/10.1016/j.cviu.2018.09.001

[10] Munmun De Choudhury, Sanket Sharma, and Emre Kiciman. 2016. Characterizing dietary choices, nutrition, and language in food deserts via social media. In *Proceedings of the 19th acm conference on computer-supported cooperative work & social computing*. 1157–1170.

[11] Yujie Dong, Adam Hoover, Jenna Scisco, and Eric Muth. 2012. A new method for measuring meal intake in humans via automated wrist motion tracking. *Applied psychophysiology and biofeedback* 37, 3 (2012), 205–215.

[12] Daniel Fried, Mihai Surdeanu, Stephen Kobourov, and Melanie Hingle. 2014. Analyzing the Language of Food on Social Media. *Proceedings - 2014 IEEE International Conference on Big Data, IEEE Big Data 2014* (09 2014). https://doi.org/10.1109/BigData.2014.7004305

[13] Jun Harashima, Yuichiro Someya, and Yohei Kikuta. 2017. Cookpad Image Dataset: An Image Collection as Infrastructure for Food Research. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR âĂŹ17)*. Association for Computing Machinery, New York, NY, USA, 1229âĂŞ1232. https://doi.org/10.1145/3077136.3080686

[14] Marvin Harris. 1998. *Good to eat: Riddles of food and culture*. Waveland Press.

[15] 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 2nd International Workshop on Multimedia Assisted Dietary Management (MADiMa âĂŹ16)*. Association for Computing Machinery, New York, NY, USA, 41âĂŞ49. https://doi.org/10.1145/2986035.2986042

[16] Christopher Holmberg, John E Chaplin, Thomas Hillman, and Christina Berg. 2016. Adolescents' presentation of food in social media: An explorative study. *Appetite* 99 (2016), 121–129.

[17] Mona Jalal, Kaihong Wang, Sankara Jefferson, Yi Zheng, Elaine O Nsoesie, and Margrit Betke. 2019. Scraping Social Media Photos Posted in Kenya and Elsewhere to Detect and Analyze Food Types. In *Proceedings of the 5th International Workshop on Multimedia Assisted Dietary Management*. 50–59.

[18] Hokuto Kagaya, Kiyoharu Aizawa, and Makoto Ogawa. 2014. Food Detection and Recognition Using Convolutional Neural Network. In *Proceedings of the 22nd ACM International Conference on Multimedia (MM âĂŹ14)*. Association for Computing Machinery, New York, NY, USA, 1085âĂŞ1088. https://doi.org/10.1145/2647868.2654970

[19] Y. Kawano and K. Yanai. 2014. Automatic Expansion of a Food Image Dataset Leveraging Existing Categories with Domain Adaptation. In *Proc. of ECCV Workshop on Transferring and Adapting Source Knowledge in Computer Vision (TASK-CV)*.

[20] 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 *International Conference on Smart Homes and Health Telematics*. Springer, 37–48.

[21] Yuzhen Lu. 2016. Food image recognition by using convolutional neural networks (CNNs). *arXiv preprint arXiv:1612.00983* (2016).

[22] Javier Marin, Aritro Biswas, Ferda Ofli, Nicholas Hynes, Amaia Salvador, Yusuf Aytar, Ingmar Weber, and Antonio Torralba. 2019. Recipe1M+: A Dataset for Learning Cross-Modal Embeddings for Cooking Recipes and Food Images. *IEEE Trans. Pattern Anal. Mach. Intell.* (2019).

[23] 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 *The IEEE International Conference on Computer Vision (ICCV)*.

[24] Vishwali Mhasawade, Anas Elghafari, Dustin T Duncan, and Rumi Chunara. 2020. Role of the Built and Online Social Environments on Expression of Dining on Instagram. *International Journal of Environmental Research and Public Health* 17, 3 (2020), 735.

[25] Marion Nestle, Rena Wing, Leann Birch, Lorelei DiSogra, Adam Drewnowski, Suzette Middleton, Madeleine Sigman-Grant, Jeffery Sobal, Mary Winston, and Christina Economos. 1998. Behavioral and Social Influences on Food Choice. *Nutrition Reviews* 56, 5 (1998), 50–64. https://doi.org/10.1111/j.1753-4887.1998.tb01732.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1753-4887.1998.tb01732.x

[26] Parisa Pouladzadeh, Shervin Shirmohammadi, and Rana Al-Maghrabi. 2014. Measuring calorie and nutrition from food image. *IEEE Transactions on Instrumentation and Measurement* 63, 8 (2014), 1947–1956.

[27] Francesco Ragusa, Valeria Tomaselli, Antonino Furnari, Sebastiano Battiato, and Giovanni M. Farinella. 2016. Food vs Non-Food Classification. In *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management (MADiMa âĂŹ16)*. Association for Computing Machinery, New York, NY, USA, 77âĂŞ81. https://doi.org/10.1145/2986035.2986041

[28] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 779–788.

[29] Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, and Silvio Savarese. 2019. Generalized Intersection over Union. (June 2019).

[30] Sanket S Sharma and Munmun De Choudhury. 2015. Measuring and characterizing nutritional information of food and ingestion content in instagram. In *Proceedings of the 24th International Conference on World Wide Web*. 115–116.

[31] 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 2nd International Workshop on Multimedia Assisted Dietary Management*. 3–11.

[32] Lone Brinkmann Sørensen, Per Møller, A Flint, Magni Martens, and A Raben. 2003. Effect of sensory perception of foods on appetite and food intake: a review of studies on humans. *International journal of obesity* 27, 10 (2003), 1152–1166.

[33] Claudia Wagner and Luca Maria Aiello. 2015. Men eat on mars, women on venus? an empirical study of food-images. In *Proceedings of the ACM Web Science Conference*. 1–3.