Scraping Social Media Photos Posted in Kenya and Elsewhere to Detect and Analyze Food Types

Mona Jalal[1*], Kaihong Wang[1*], Sankara Jefferson[2], Yi Zheng[1], Elaine O. Nsoesie[1], Margrit Betke[1]
[1]Boston University, [2] Lori Systems LTD

ABSTRACT

Monitoring population-level changes in diet could be useful for education and for implementing interventions to improve health. Research has shown that data from social media sources can be used for monitoring dietary behavior. We propose a **scrape-by-location** methodology to create food image datasets from Instagram posts. We used it to collect 3.56 million images over a period of 20 days in March 2019. We also propose a **scrape-by-keywords** methodology and used it to scrape ~30,000 images and their captions of 38 Kenyan food types. We publish two datasets of 104,000 and 8,174 image/caption pairs, respectively. With the first dataset, **Kenya104K**, we train a **Kenyan F**ood **C**lassifier, called **KenyanFC**, to distinguish Kenyan food from non-food images posted in Kenya. We used the second dataset, **KenyanFood13**, to train a classifier **KenyanFTR**, short for **Kenyan F**ood **T**ype **R**ecognizer, to recognize 13 popular food types in Kenya. The KenyanFTR is a multimodal deep neural network that can identify 13 types of Kenyan foods using both images and their corresponding captions. Experiments show that the average top-1 accuracy of KenyanFC is 99% over 10,400 tested Instagram images and of KenyanFTR is 81% over 8,174 tested data points. Ablation studies show that three of the 13 food types are particularly difficult to categorize based on image content only and that adding analysis of captions to the image analysis yields a classifier that is 9 percent points more accurate than a classifier that relies only on images. Our food trend analysis revealed that cakes and roasted meats were the most popular foods in photographs on Instagram in Kenya in March 2019.

KEYWORDS

Image datasets, food detection, data mining, social media

ACM Reference Format:
Mona Jalal, Kaihong Wang, Sankara Jefferson, Yi Zheng, Elaine O. Nsoesie, Margrit Betke. 2019. Scraping Social Media Photos Posted in Kenya and Elsewhere to Detect and Analyze Food Types. In *5th International Workshop on Multimedia Assisted Dietary Management (MADiMa'19), October 21, 2019, Nice, France.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3347448.3357170

*The first and second authors contributed equally to this research.

MADiMa '19, October 21, 2019, Nice, France

ACM ISBN 978-1-4503-6916-9/19/10…$15.00
https://doi.org/10.1145/3347448.3357170

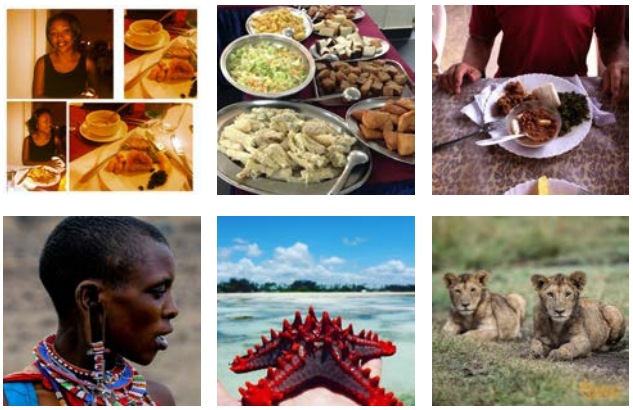

Figure 1: Sample images of the proposed Kenya104K dataset.

1 INTRODUCTION

Data from Instagram can be used to study a community's food consumption patterns as people post images of what and where they eat. Instagram users share their activities, moods, and location of particular venues in real time, thereby allowing for aspects of their experiences and patterns of eating and drinking to be captured. This enables the exploration of questions such as, why do people share images on social media and what types of foods do people tend to share on social media [9, 37]. Instagram hashtags have been shown to be useful in the study of the characteristics of food as well as the context and interests of users related to food [31]. Images and hashtags together can provide insights into how particularly youths perceive and interact with food in their communities, allowing researchers to study the diverse aspects of food culture in a specific demographical group. Furthermore, data on food and location from Instagram can allow researchers to establish eating patterns and their possible associations to conditions like obesity [2, 38].

Our work is unique in its focus on an African country (Fig. 1). There is a dearth of studies evaluating the use of social media for studying nutrition and diet in African countries using computer vision algorithms. We take advantage of the wide adoption of mobile technology and social media in African countries, especially in urban areas to provide tools that enable social scientists to assess the utility of social media data for studying diets and attitudes towards foods in Kenya. Specifically, in this paper, we describe methods to produce large-scale datasets that can be used to study food trends on Instagram. We then apply these methods to collect and analyze datasets of Kenyan food images and their captions.

Kenya has a dietary culture that is distinct from Western countries, so we could not simply apply previously developed food type detectors on Kenyan Instagram posts. In order to develop our own

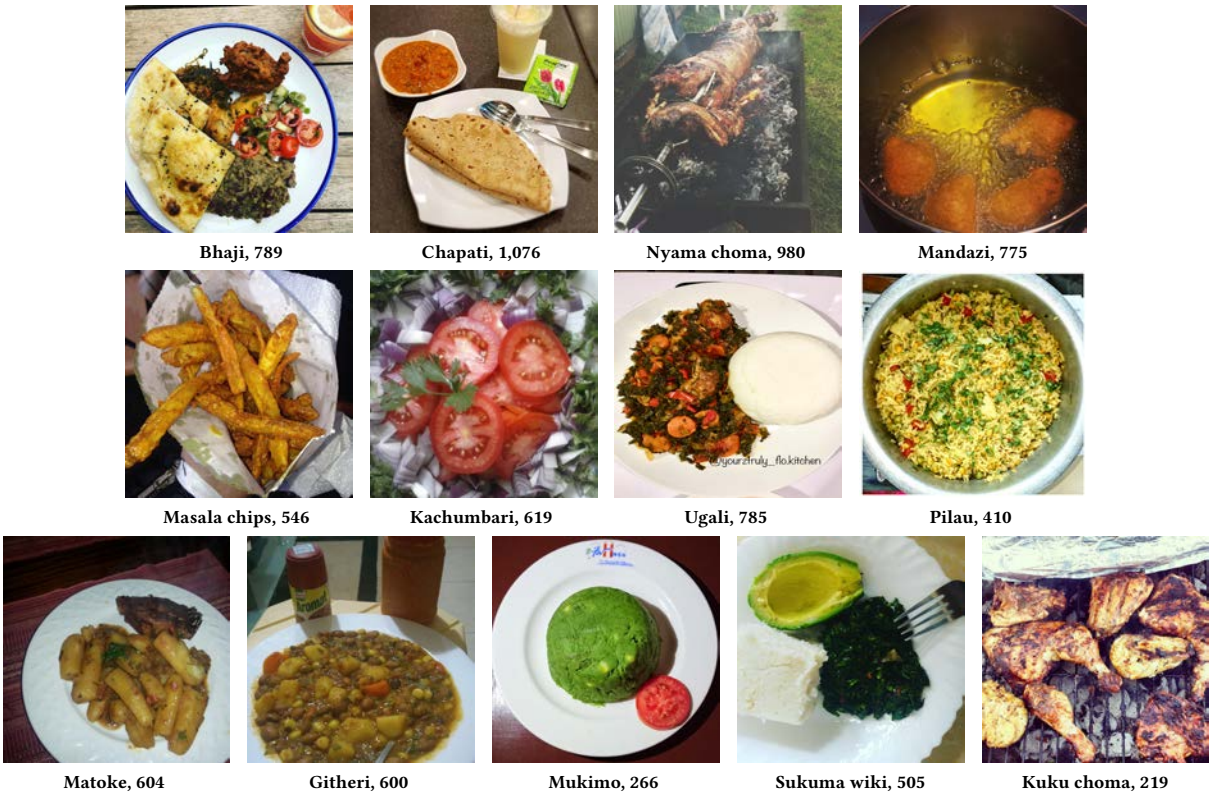

Figure 2: Sample images of Kenyan food types in the proposed KenyanFood13 dataset with numbers of samples for each type.

Kenyan food type detector, we had to overcome the challenge of automating the scraping and collecting of training images with labels that define the food type. Another challenge was to effectively scrape and filter extremely large numbers of Instagram images from a specific geographic region. Since we downloaded all Kenya-originated images from Instagram during a twenty day period in March 2019, we can use our classifier (as well as existing classifiers for foods not specific to Kenya) to analyze the prevalence of food images as well as food types (Fig. 2) in Kenyan Instagram posts during that period.

As far as we know, there is no publicly available food image dataset dedicated to the task of Kenyan food recognition. Our efforts to create Kenyan-specific datasets and classifiers are crucial for any food trend analysis a health informaticist may want to conduct for Kenya. Dibia [12] and Buolamwini and Gebru [4] warned that the generalization of learning systems is undermined when they are trained on datasets without African content or context in African scenarios. Our results show, while food-detection classifiers trained on food image datasets that only consist of images of Western or East Asian foods can recognize whether food is shown in an image relatively accurately, results improve when a classifier is trained specifically on Kenyan images. For the task of food type recognition, any supervised learning system needs to have examples of specific Kenyan foods to be able to recognize them (and food like "sukuma wiki" is not represented in previous datasets).

To summarize, our contributions are as follows:

- We propose two systems for collecting Instagram posts, a **Scrape-by-Location** system and a **Scrape-by-Keywords** system to collect posts about Kenyan foods uploaded in Kenya and elsewhere.
- We used our scraping systems to collect millions of Instagram images and their captions, and then devised methods to filter these posts in order to create two datasets. Our multimodal dataset, **Kenya104K**, can be applied to train machine learning systems to detect Kenyan foods in images. Our Kenyan food type dataset, **KenyanFood13**, contains 13 types of popular and representative Kenyan foods.
- We propose a **multimodal deep learning model, KenyanFTR,** that interprets feature vectors, which in combination represent food images and their corresponding captions, to predict the type of Kenyan food shown in an image. We also make a classifier, **KenyanFC**, available that is trained to distinguish food (both Kenyan and non-specific food) from non-food images.
- We applied our techniques to the millions of Instagram posts we collected across Kenya over a period of 20 days to give an example of the kind of analysis social scientists may conduct with our tools.

Our datasets, code and archival version paper will be publicly available [1]. We hope that our work encourages other researchers to include images from Africa in their image analysis research by either following our methodology to create their own Instagram datasets or using our Kenyan image data to further develop computer vision tools to recognize and study image content.

2 RELATED WORK

Studies have shown that image and text data from social media sites, such as Instagram and Twitter, can be useful for monitoring diet [16, 32, 42] or identifying deserts [9]. Images of food have been used in apps to measure food calories in real time (e.g., https://caloriemama.ai/) and to log food [28, 48]. Various datasets have been collected from social media (e.g., [5, 34, 35, 41, 49]).

2.1 Food/Non-food Datasets and Classification

Datasets built to support food detection tasks have two classes, namely "food" and "non-food." Images in the food class include various kinds of foods while images in the non-food class should cover as many other objects as possible (human portraits, landscape scenes, and other objects) that could appear in use-case datasets in which food is supposed to be detected. Table 1 lists some popular food/non-food datasets as well as our proposed dataset Kenya104K.

Table 1: Food/Non-food datasets

Dataset	# food images	# non-food im.	total
Food-5K [43]	2,500	2,500	5,000
IFD [21]	4,230	5,428	9,658
FCD [21]	25,250	28,322	53,572
Flickr-Food [15]	4,805	0	4,805
Flickr-NonFood [15]	0	8,005	8,005
Kenya104K	52,000	52,000	104,000

Many methods and models have been proposed for the task of food/non-food classification. Early work in analyzing food images by Kitamura et al. [24] used hand-crafted features such as color histograms, Discrete Cosine Transform coefficients, and shape representations to train a support vector machine (SVM) classifier. Farinella et al. [15] used a one-class SVM on features extracted based on a bag-of-words approach to exclude the influence of non-food images. More recent research efforts have used deep networks for food detection and shown their superior performance over traditional approaches (e.g., [22, 43] applied *GoogLeNet* [45] and *Network in Network* [26]). Ragusa et al. [39] investigated how to optimally combine different representation methods with classification schemes.

2.2 Food Type Datasets and Classification

The task of recognizing food types in images is typically solved with a supervised learning model and under the assumption that images are known to contain food. For training food classification models, researchers have proposed many food type datasets collected in various ways. For example, Bossard et al. [3] built the ETHZ Food-101 dataset by collecting food images from http://foodspotting.com and randomly sampling 1,000 images from the top 101 most popular dishes with consistent names ranked on the website. Table 2 lists some popular food/non-food datasets, as well as our proposed dataset KenyanFood13.

[1] https://github.com/monajalal/Kenyan-Food

Table 2: Food type datasets

Dataset	# classes	# img. per class	Total # of images	Style of food
ETHZ Food-101 [3]	101	1,000	101,000	As, E, Am[1]
UPMC Food-101 [49]	101	1,000	101,000	As, E, Am
UEC-FOOD-100 [30]	100	~90	9,060	Japanese
UEC-FOOD-256 [23]	256	~127	31,397	Japanese
VireoFood-172 [5]	172	~641	110,241	Chinese
UNICT-FD889 [14]	889	~4	3,583	As, E, Am
UNICT-FD1200 [13]	1200	~4	4,754	As, E, Am
Food-524DB [8]	524	~473	247,636	As, E, Am
PFID [6]	101	18	1,818	E, Am
Food500 [33]	500	~300	150,000	As, E, Am
NTU-FOOD [7]	50	100	5,000	Chinese
KenyanFood13	13	~629	8,174	Kenyan

1 Asian (As), European (E), American (Am)

Classic food type recognition models generally follow the pipeline of extracting and combining different features and feeding them into a classifier (e.g., a SVM). For example, Joutou and Yanai [20] trained a multiple kernel SVM using a combination of features, including Gabor texture features and color histograms. Bossard et al. [3] proposed a method that applies random forests to extract discriminative visual components from the ETHZ Food-101 dataset. Deep neural networks have shown exciting performance in food recognition tasks. Bossard et al. [3] showed that an AlexNet [25] trained on the ETHZ Food-101 dataset can achieve higher accuracy than other methods they tried. Yanai and Kawano [51] showed the effectiveness of fine-tuning a pre-trained AlexNet for food recognition in images. Martinel et al. [29] applied residual learning to the food recognition task by introducing a "wide-slice residual network."

3 METHODS

In this section, we first describe two multimodal dataset collection methods, then the resulting datasets, and finally the classifiers we designed to analyze them. We developed the "scrape-by-location" method to collect images on Instagram from Kenya and the "scrape-by-keywords" method to collect popular Kenyan food images posted on Instagram, but not necessarily from Kenya. Our systems are shown in Figures 3 and 4.

3.1 Scrape-by-Keywords

Our scrape-by-keywords data collection method relied on Kenyan experts to provide us with a list of foods that are popular in Kenya. We received a list of 38 food names in the Kiswahili language. We used the keyword searching module of the Instagram Scraping API [1] to search for Instagram posts that included at least one of the 38 food names in their image captions (usually tagged with a hashtag) and downloaded both images and captions. The API provided us with ~40,000 data points and stopped finding additional posts after one day of search (March 23–24, 2019). We manually inspected and filtered out about 10,000 images that did not include any food. Our scrape-by-keywords process thus resulted in about

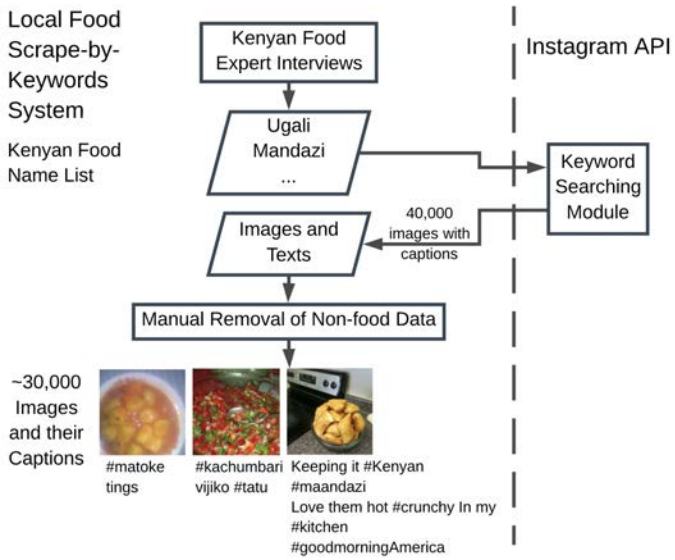

Figure 3: Overview of the proposed Scrape-by-Keywords System.

30,000 Kenyan food images and their captions, which include the names of the foods in Kiswahili.

3.2 Scrape-by-Location

We defined a set of rectangular regions on the map of Kenya (red boxes in Fig. 4) to cover its territory and then defined a grid within these regions. The grid has a stride of 0.02 degree of longitude and 0.02 degree of latitude. For each point on the grid we searched all nearby locations registered on Instagram using the location searching module of the Instagram Scraping API. We recorded the identifier (ID) of each retrieved location. Using the ID, we then applied the Instagram post searching module of the API to retrieve recent posts that had been uploaded from the location with that ID. For each post, we downloaded as much information as available. This included the primary key, ID, image(s), URL(s), and potentially caption, and latitude and longitude of the location. We applied our scrape-by-location system for 20 days, spanning March 7 to March 27, 2019, and retrieved 2.7 million Instagram posts. The posts contained a total of 3.56 million images, many with captions.

3.3 Kenyan Food Type Dataset *KenyanFood13*

To develop a food type dataset, we analyzed ~30,000 images that we collected with our scrape-by-keywords system. We noticed that the number of images per food type differed significantly. We collected 9,142 images of bhaji, but only 316 images of mukimo. Only 15 of the 38 food types were represented by more than 500 images. Moreover, there is no guarantee that the images downloaded by the scrape-by-keywords system actually include the types of food that their captions mention. All we know is that the captions include at least one of the 40 Kenyan food names that we used as keywords in the collection process. Considering these issues, we decided to reduce the dataset to only 13 classes, where each class has at least 200 samples. We manually inspected each remaining image to ensure that it is a photograph of the food type that it was assigned to. This

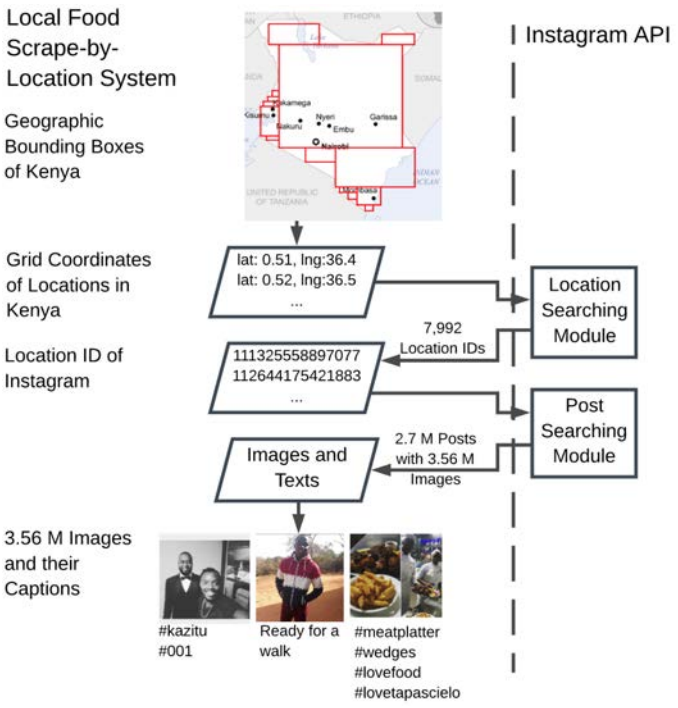

Figure 4: Overview of the proposed Scrape-by-Location System

process resulted in 8,174 images in 13 food type classes. Sample images of our KenyanFood13 dataset and the number of images in each of the 13 food type classes are shown in Fig. 2.

3.4 *Kenya104K* Dataset

In order to discover eating patterns based on analyzing large datasets (like the 3.5 million multimodal dataset we collected with our scrape-by-location method), health informaticists need a tool to distinguish images of food from non-food. To create such a tool, a classifier needs to be trained on images that are categorized into food and non-food. To create a dataset automatically (versus by crowdsourcing), a classifier is needed. To resolve this causality dilemma, we build up our training set in piecemeal. We first assign the 30,000 images retrieved by our scrape-by-keywords system to the class of food images. We also use 30,000 non-food images from our scrape-by-location collection, which we selected manually. We then combine them with 9,658 images from the Instagram Food Dataset (IFD) and 53,572 images from FCD (food and non-food). Then, we trained a food/non-food detector that detects the food images collected from Kenya. By applying the trained food detector, we initially detected about 70,000 food images and after manual inspection, we ended up with 52,000 food images. To create negative samples, we added 52,000 manually inspected random samples from the images collected by "scraping-by-location."

3.5 Our Classifier *KenyanFC*

To develop our classifier **KenyanFC**, which can distinguish between food and non-food content in images, we fine-tuned ResNeXt101 [50], pre-trained on ImageNet dataset [10], with a

merged dataset containing FCD [21], Food-5K [43], as well as our own food/non-food dataset KenyanFood104k. The number of nodes of the output layer of ResNeXt101 was changed to two to adapt to the food/non-food classification task.

3.6 Our Classifier *KenyanFTR*

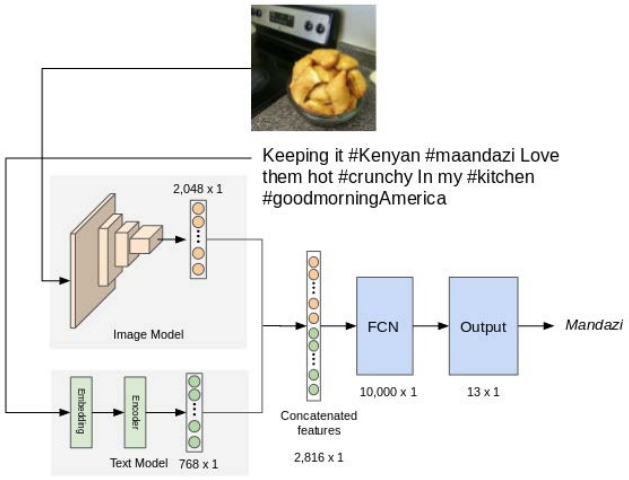

Figure 5: Architecture of food type recognition model (FCN stands for fully connected network).

To develop a classifier that can distinguish among 13 Kenyan food types, we propose a feature fusion model that extracts features from each image and its corresponding caption. The system architecture of our model KenyanFTR is shown in Figure 5.

The model uses the BERT language model [11] to extract features from the Instagram image captions and fuses these features with the extracted image features. We passed each input token –the words in the caption– through three embedding layers (token embedding, segment embedding, and position embedding) and these three embedding representations are summed element-wise to produce a single input representation. The input representation is passed to the encoder layer of BERT, and we used the outputs as features representing text. To represent the image, we applied the ResNeXt101 model pre-trained on ImageNet and extracted a feature vector from the last hidden layer. After we extracted features from both BERT and ResNeXt101, we concatenated the two feature vectors into a new vector followed by a hidden layer with 10,000 neurons and an output layer with 13 classes.

4 EXPERIMENTAL RESULTS

We first introduce experiments with our classifier KenyanFC on our food/non-food dataset Kenya104K followed by experiments with our multimodal model KenyanFTR on the KenyanFood13 dataset. Finally, we apply our KenyanFTR model and other methods to thousands of Instagram posts to analyze food trends in Kenya.

4.1 Experiments for Food/Non-food Detection

We first fine-tuned ResNeXt101 (pre-trained on ImageNet) with either Food-5K, FCD, or Kenya104K. For all datasets, 90% of the images were used for training and validation (72% train, 18% validation) in five folds (five different splits). The resulting five models were tested on a hold-out test set of 10% of the dataset. Accuracy mean and standard deviations were then reported as averages of these five models (Table 3). To train our KenyanFC, we built a training set by merging the training, validation and testing set of Food-5K and FCD with the training and validation set of Kenya104K. Finally KenyanFC was evaluated on the testing set of Kenya104K. During the training phase, we applied data augmentation including random rotation, horizontal flipping, and color jitter. During the fine-tuning process, we used stochastic gradient descent (SGD) as our optimizer with a learning rate of 0.0001 and a momentum of 0.9. To avoid over-fitting, we chose the model producing the highest validation accuracy as the final model within 10 epochs of the fine-tuning process. Our results show high accuracy in food detection (above 95% in all cases shown in Table 3). The testing accuracy of our KenyanFC (99%) shows that it generalized better when it evaluated images in the Kenyan context.

Table 3: Accuracy of food/non-food classification. Table entry (i, j) means that the classifier was trained on dataset i and tested on dataset j.

Dataset	Food-5K	FCD	Kenya104K
Food-5K	99.24%± 0.08%	98.54% ± 0.12%	95.20% ± 0.91%
FCD	98.44%± 0.45%	99.52% ± 0.02%	95.90% ± 0.39%
Kenya104K	98.32%± 0.20%	97.89%± 0.37%	99.03%± 0.06%
Combined	N/A	N/A	99.01%± 0.06%

4.2 Experiments for Food Type Recognition

Our classifier KenyanFTR yields a top-1 accuracy of almost 81% when tested in a 5-fold cross-validation manner on KenyanFood13. For each class, most predicted labels match the ground truth labels, as can be seen in the confusion matrix in Fig. 6. For six classes, recognition accuracy surpassed 90%. For comparison, we also present the confusion matrix for the classifier that only interprets images.

We note that, by design, the scrape-by-keywords data-collection method only harvests images with captions that contain the names of the Kenyan foods we targeted in our scraping. For such images, to identify the food type, a classifier would not even have to analyze the image and could just evaluate its caption. We found that a pre-trained BERT model (we tested the uncased version) reaches an accuracy level of about 98% in identifying the food type from the captions alone. This means the model can almost always make correct predictions when food names are available. It would be wrong, however, to assume that any Kenyan who posts food images on Instagram also includes food names in their captions. In fact, we found that for only 1,914 images out of the 52,000 images that we obtained with our scrape-by-location method, the captions contain food names. To train our KenyanFTR classifier to identify a food type in an image without relying on the food type name appearing in the caption, we removed any hashtagged food names from the captions in our images before using them in our experiments.

In order to explore the value of taking advantage of the two modalities, image and text, in our KenyanFood13 dataset, we conducted ablation studies with models that take as input only images or only text (Table 4). For the former, we fine-tuned a ResNeXt101

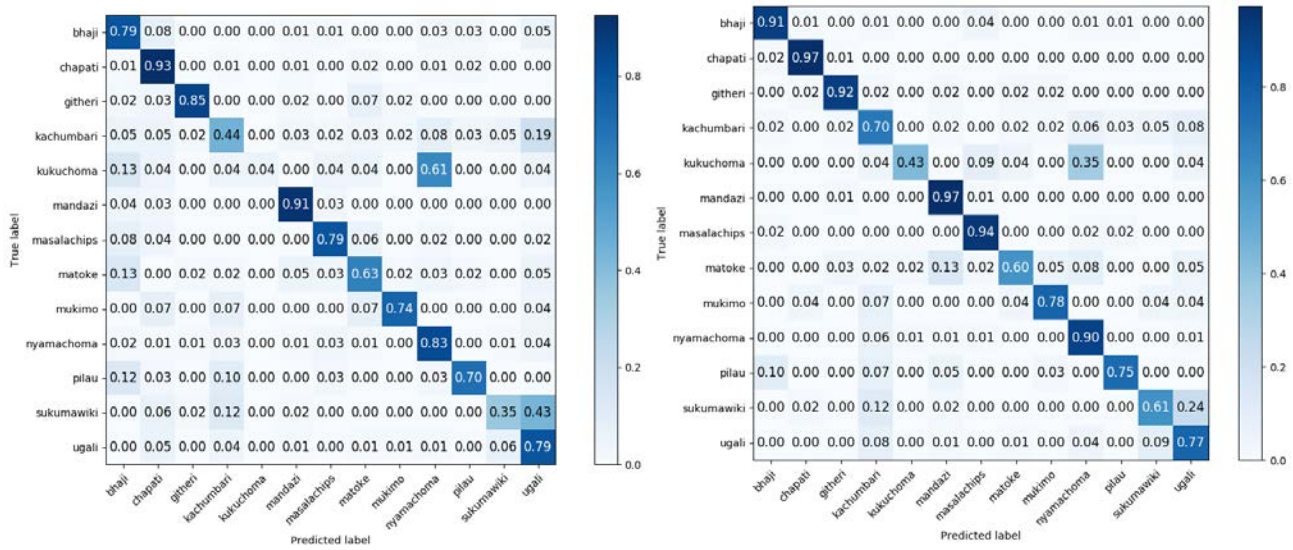

Figure 6: KenyanFTR applied to KenyanFood13: Confusion matrices of food type recognition model based on only images (left) and based on images and captions (right).

Table 4: Results of Ablation Studies: Accuracy of different input settings on KenyanFood13.

Method	Test Accuracy	
	Top-1	Top-3
Image only	73.18%± 0.79%	92.04%± 0.44%
Caption only	65.30%± 1.70%	83.68%± 1.55%
Ours: Image + Caption	81.04%± 0.86%	95.95%± 0.44%

Table 5: Results of Comparison Experiments: Accuracy of different models on KenyanFood13.

Method	Test Accuracy	
	Top-1	Top-3
InceptionV3+BERT	71.92%± 1.52%	88.57%± 0.68%
InceptionV4+BERT	67.40%± 1.49%	85.05%± 1.93%
ResNet101+BERT	76.74%± 2.02%	93.71%± 1.18%
DenseNet161+BERT	79.02%± 0.96%	95.14%± 0.73%
Ours: ResNeXt101+BERT	81.04%± 0.86%	95.95%± 0.44%

(pre-trained on ImageNet) only with the images of KenyanFood13 and evaluated its performance, while for the latter, we fine-tuned a pre-trained BERT-based model using only the captions of KenyanFood13. Finally, we compared their performance with our KenyanFTR model, which takes both images and text as input. The gain of accuracy in using both image and text modalities is significant – the additional use of text improves the top-1 accuracy result by more than 9 percent points, while the additional use of the image improves the top-1 accuracy by 30 percent points.

We also investigated the performance of different image feature extractors. We compared the ResNeXt101 feature extractor used by our KenyanFTR with other popular pre-trained deep learning models, including ResNet101 [17], InceptionV3 [46], InceptionV4 [44], and DenseNet161 [18]. We note that the feature vectors computed by ResNet101, InceptionV3, InceptionV4, and DenseNet161 have lengths 2,048, 2,048, 1,536, and 2,208, respectively.

During the fine-tuning process in the experiment, we applied 5-fold cross-validation to train and test our KenyanFood13 dataset. Further, we applied the same data augmentations as in our food/no-food classifier (random rotation, flipping, color jitter). During the fine-tuning process, we used SGD as the optimizer with a learning rate of 0.0001, momentum of 0.9, and 12 epochs. The top-1 testing accuracy of the different models is below 80% except for our KenyanFTR, which achieves an accuracy of 81% (Table 5).

4.3 Analysis of Food Trends in Kenya

To investigate food preferences in Kenya (i.e., Instagram sharing preferences), we designed a tool to recognize food types in the 52,000 food images of the Kenya104K dataset. Considering the fact that we conducted an intensive grid search within the geographic boundaries of Kenya, which ensured broad coverage during the collecting process, we suggest that our dataset is sufficiently representative for research of food trends on Instagram in Kenya.

To identify food type images uploaded in Kenya, we first applied our KenyanFTR model on the 52,000 food images of the Kenya104K dataset. Because KenyanFTR was trained on KenyanFood13, which includes data of 13 popular Kenyan foods, we reason that a confidence score of at least 70% of a food type label predicted by KenyanFTR on these images is likely correct. By visual inspection of the food images of Kenya104K, we also noticed that fruits and some Western foods such as cake and pizza are popular in Kenyan Instagram uploads. We therefore applied YOLO v3 [40], pre-trained on MSCOCO [27], to detect fruits and Western foods in the food images of Kenya104K. We report that 25,865 images were predicted to depict food. This includes 13,975 images with Kenyan food items, 2,530 images with fruits and vegetables, and 13,860 with Western foods (details in Table 6). Note that for a single image, the classifiers

Table 6: Some of the food types detected in Kenyan Instagram posts collected during 20 days in March 2019.

Food type	# images	Food type	# images
cake	7,559	nyamachoma	3,220
kachumbari	2,990	mandazi	2,671
pizza	2,456	sandwich	2,407
masalachips	1,315	githeri	894
doughnut	851	pilau	728
carrot	702	banana	628

may yield multiple prediction results since the images may indeed contain more than one type of food.

5 DISCUSSION OF RESULTS

Torralba and Efros [47] discussed the concept of dataset bias – it is inevitable that datasets have intrinsic features that may be difficult to recognize by humans, for example, due to selection, image capture, or category-preference biases. We suggest that this could be studied as "a feature not a bug" for our datasets. The high occurrence of the word "love" in image captions of Kenya104K , for example, seems to indicate that the Instagram users prefer to upload images of foods that they love (for additional examples, see Fig. 7). A capture bias is most likely present in our data since people typically center objects of interest in a photograph. The range of numbers of images per each food type that we collect indicates a category-preference bias. Other challenges are the existence of multiple food items in the images, image quality (low resolution or out-of-focus), as well as unknown food types.

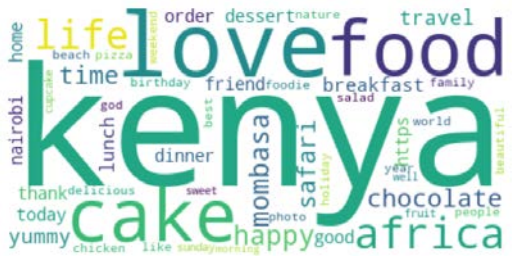

Figure 7: The word cloud shows words in captions of the 52K food images in Kenya104K, resized proportionally to how frequently they occurred.

5.1 Discussion of Food/Non-food Detection

Supervised learning systems typically perform better when training and testing images belong to subsets of the same dataset. This is also true for our classifier *KenyanFC*, as can be seen in Table 3. From this table, we observe that, when tested on Kenya104K, models trained on Food-5K and FCD performed somewhat worse than a model trained on only Kenya104K, or a model trained on the combination of all three datasets. Also, the difference between accuracy of models trained on Kenya104K and the combination of datasets is not statistically significant (i.e., the difference in mean accuracy is smaller than the standard deviations in accuracy).

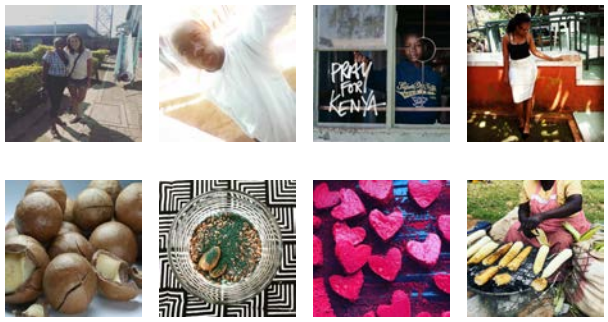

Figure 8: Images correctly classified by KenyanFC but misclassified by ResNeXt101 when trained on FCD and tested on our Kenya104K. Images in the first row are non-food images misclassified as food images; images in the second row are food images misclassified as non-food images.

To explore the advantage of Kenya104K over other food/non-food datasets when processing images from Kenya, in Figure 8, we provide some example images that were misclassified by the model trained on FCD but correctly classified by the model trained on Kenya104K. It is obvious that images containing Kenyan content are confusing to the models since they are trained on other datasets that do not contain such content. This illustrates the necessity of having a dataset dedicated to a specific country or region for image recognition tasks when trying to analyze food trends in Africa.

5.2 Discussion of Food Type Recognition

Our model comparison experiments reveal that all five deep models fused with BERT generalized well on our dataset, with KenyanFRT performing the best (Table 5). For KenyanFTR, we studied the remaining challenges as follows. For every image in a selected class, we computed their L2 distance to all other images not belonging to the current food type and found the pair of images with the smallest distance. Two examples of such similar image pairs are shown in Figure 9.

Figure 9: Examples of most confused food images in our KenyanFood13.

Our ablation study shows that the difference in accuracy of three models given images only, caption only, and images combined with captions is large (Table 4). The model trained on images achieved a top-1 accuracy of 73.18% and top-3 accuracy of 92.04%, while the model trained on captions achieved top-1 accuracy of 65.30% and top-3 accuracy of 83.68%, reflecting the fact that images are more informative when training models. However, when we combined images with their corresponding captions, recognition accuracy increased from 73% to 81%, which reveals the advantage of multi-modality of our KenyanFood13 dataset. Interestingly, comparing

the confusion matrices (Fig. 6), we observe that the information brought by captions helped significantly in reducing the occurrence of misclassification. For example, "kachumbari" and "sukuma wiki" are very likely to be recognized as "ugali" since they are always served and eaten with "ugali," and so they frequently appear together in the images. After including the captions into the training process, the misclassification rate of "kachumbari" and "sukuma wiki" with respect to "ugali" decreased from 19% and 43% to 8% and 24%, respectively. Also, the misclassification rate between "kuku choma" and "nyama choma" was significant simply because they look similar, but after including the captions along with the images, the misclassification rate dropped significantly from 61% to 35%.

5.3 Discussion of Food Trend Analysis

A healthy diet is important for good health and can protect against diseases such as diabetes and cardiovascular disease [36]. According to the World Health Organization, healthy diets for adults should be high in vegetables and fruits, have less than 30% of total energy consumption from fat, and have less than 5 g of salt content per day. Access to a healthy diet, however, can be hindered by factors such as inability to afford healthy foods and limited access to healthy food options [19, 52].

The relationship between socioeconomic status, diet, and obesity is not always straightforward. For example, the risk for obesity can increase with wealth, as was shown in a study conducted in the Karonga District and Lilongwe city in Malawi [38]. In contrast, several studies have shown positive associations between lower socioeconomic status and lack of access to healthy foods and higher obesity prevalence [2]. Understanding changes in attitudes and sentiments towards unhealthy foods can be useful for education and implementation of interventions to improve health in communities.

A map that reveals the foods most popular for Instagram uploads in each county in Kenya is shown in Fig. 10. It is based on our analysis of the 52,000 food images of Kenya104K. Cake is the most popular food type for Instagram uploads in many counties of Kenya, as well as developed areas such as Nairobi and Mombasa. A likely reason is that people enjoy sharing images of cakes on social media, especially during celebrations (e.g., birthday cakes). Interestingly, other Western foods, such as pizza and sandwiches, are popular in some remote areas, while people living in south-central Kenya prefer to upload images of classic Kenyan foods such as "nyama choma"and "mandazi."

To take a more intuitive look at the 52,000 food images, we created a word cloud of their captions, which is shown in Figure 7. High frequency of words like "travel," "Kenya," "Africa" may indicate that many of these posts are sent by tourists, since these words are not likely to be used by local people living in the country.

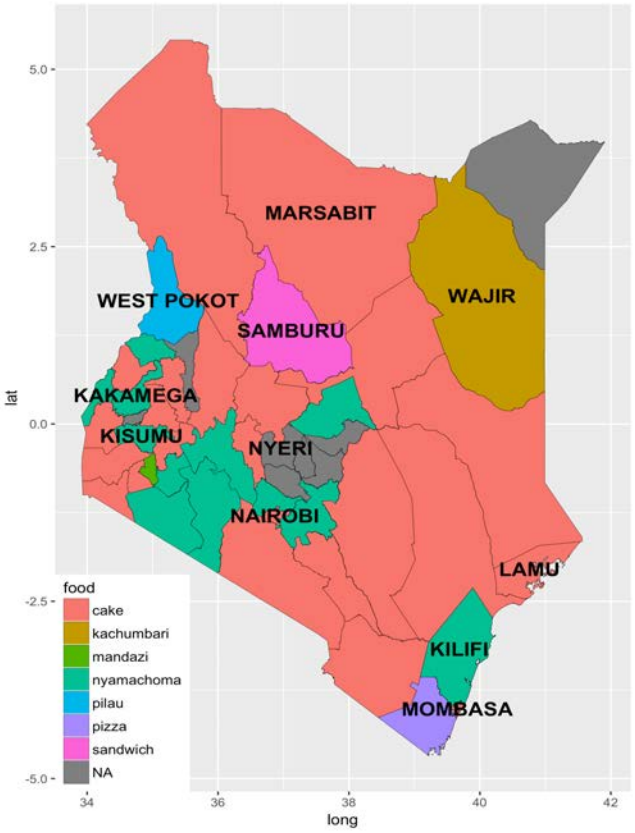

Figure 10: Most popular types of food per county in Kenya, according to 20 days of Instragram posts in March 2019.

6 CONCLUSIONS AND FUTURE WORK

In this paper, we presented two systems to scrape social media photos and their associated metadata, scrape-by-keywords and scrape-by-location, and two datasets that we developed with the help of these systems, Kenya104K for the food/non-food detection task, and KenyanFood13 for food type recognition tasks. Extensive experiments revealed the advantages of having a Kenyan-food-specific dataset for training a classifier to detect such food and of having a multimodal dataset for classification of Kenyan food types. We note that our food type classifier is Instagram-agnostic and can be applied to images with or without captions. We applied the classifiers, in combination with existing food-image and text classifiers, to 3.56 million images that were posted on Instagram across Kenya over a period of 20 days. As an example of the health science analysis that our work enables, we reveal that the most popular foods for Instagram uploads in Kenya were cake and roasted meat. Social scientists may use our datasets and/or our data scraping processes and classifiers to analyze food trends, dietary values, geographical differences, the impact of tourism, etc., by collecting Instragram posts over additional periods of time.

Our code and datasets are publicly available and could be augmented by annotations such as cuisine type, flavors, or ingredients. Furthermore, a mobile phone application could be devised that would inform Kenyan users, in real time, of the dietary values of the meals they are eating or are interested in eating.

Acknowledgements. The authors thank the National Science Foundation (1838193) and the Hariri Institute for Computing and Computational Science & Engineering at Boston University for partial support of this work.

REFERENCES

[1] Instagram API. 2019. Instagram API, Last accessed on April, 2019. https://github.com/mgp25/Instagram-API

[2] Troy C Blanchard and Todd L Matthews. 2007. Retail Concentration, Food Deserts, and Food-disadvantaged Communities in Rural America. *Remaking the North American food system: Strategies for sustainability* (2007), 201–215.

[3] Lukas Bossard, Matthieu Guillaumin, and Luc J. Van Gool. 2014. Food-101 - Mining Discriminative Components with Random Forests. In *Computer Vision - ECCV 2014 - 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VI*. 446–461. https://doi.org/10.1007/978-3-319-10599-4_29

[4] Joy Buolamwini and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Conference on Fairness, Accountability and Transparency, FAT 2018, 23-24 February 2018, New York, NY, USA*. 77–91. http://proceedings.mlr.press/v81/buolamwini18a.html

[5] Jingjing Chen and Chong-Wah Ngo. 2016. Deep-based Ingredient Recognition for Cooking Recipe Retrieval. In *Proceedings of the 2016 ACM Conference on Multimedia Conference, MM 2016, Amsterdam, The Netherlands, October 15-19, 2016*. 32–41. https://doi.org/10.1145/2964284.2964315

[6] Mei Chen, Kapil Dhingra, Wen Wu, Lei Yang, Rahul Sukthankar, and Jie Yang. 2009. PFID: Pittsburgh Fast-food Image Dataset. In *2009 16th IEEE International Conference on Image Processing (ICIP)*. IEEE, 289–292.

[7] Mei-Yun Chen, Yung-Hsiang Yang, Chia-Ju Ho, Shih-Han Wang, Shane-Ming Liu, Eugene Chang, Che-Hua Yeh, and Ming Ouhyoung. 2012. Automatic chinese food identification and quantity estimation. In *SIGGRAPH Asia 2012 Technical Briefs*. ACM, 29.

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

[9] 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*. ACM, 1157–1170.

[10] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. Ieee, 248–255.

[11] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*. 4171–4186. https://aclweb.org/anthology/papers/N/N19/N19-1423/

[12] Victor Dibia. [n. d.]. COCO-AFRICA: ACuration TOOL AND DATASET OF COMMON OBJECTS IN THE CONTEXT OF AFRICA. ([n. d.]).

[13] Giovanni Maria Farinella, Dario Allegra, Marco Moltisanti, Filippo Stanco, and Sebastiano Battiato. 2016. Retrieval and Classification of Food Images. *Comp. in Bio. and Med.* 77 (2016), 23–39. https://doi.org/10.1016/j.compbiomed.2016.07.006

[14] Giovanni Maria Farinella, Dario Allegra, and Filippo Stanco. 2014. A Benchmark Dataset to Study the Representation of Food Images. In *Computer Vision - ECCV 2014 Workshops - Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part III*. 584–599. https://doi.org/10.1007/978-3-319-16199-0_41

[15] Giovanni Maria Farinella, Dario Allegra, Filippo Stanco, and Sebastiano Battiato. 2015. On the Exploitation of One Class Classification to Distinguish Food Vs Non-Food Images. In *New Trends in Image Analysis and Processing - ICIAP 2015 Workshops - ICIAP 2015 International Workshops: BioFor, CTMR, RHEUMA, ISCA, MADiMa, SBMI, and QoEM, Genoa, Italy, September 7-8, 2015, Proceedings*. 375–383. https://doi.org/10.1007/978-3-319-23222-5_46

[16] Daniel Fried, Mihai Surdeanu, Stephen Kobourov, Melanie Hingle, and Dane Bell. 2014. Analyzing the Language of Food on Social Media. In *2014 IEEE International Conference on Big Data (Big Data)*. IEEE, 778–783.

[17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*. 770–778. https://doi.org/10.1109/CVPR.2016.90

[18] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. 2017. Densely Connected Convolutional Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*. 2261–2269. https://doi.org/10.1109/CVPR.2017.243

[19] Ehimario U Igumbor, David Sanders, Thandi R Puoane, Lungiswa Tsolekile, Cassandra Schwarz, Christopher Purdy, Rina Swart, Solange Durão, and Corinna Hawkes. 2012. "Big Food," The Consumer Food Environment, Health, and the Policy Response in South Africa. *PLoS medicine* 9, 7 (2012), e1001253.

[20] Taichi Joutou and Keiji Yanai. 2009. A Food Image Recognition System with Multiple Kernel Learning. In *Proceedings of the International Conference on Image Processing, ICIP 2009, 7-10 November 2009, Cairo, Egypt*. 285–288. https://doi.org/10.1109/ICIP.2009.5413400

[21] Hokuto Kagaya and Kiyoharu Aizawa. 2015. Highly Accurate Food/Non-Food Image Classification Based on a Deep Convolutional Neural Network. In *New Trends in Image Analysis and Processing - ICIAP 2015 Workshops - ICIAP 2015 International Workshops: BioFor, CTMR, RHEUMA, ISCA, MADiMa, SBMI, and QoEM, Genoa, Italy, September 7-8, 2015, Proceedings*. 350–357. https://doi.org/10.1007/978-3-319-23222-5_43

[22] Hokuto Kagaya, Kiyoharu Aizawa, and Makoto Ogawa. 2014. Food Detection and Recognition using Convolutional Neural Network. In *Proceedings of the ACM International Conference on Multimedia, MM '14, Orlando, FL, USA, November 03 - 07, 2014*. 1085–1088. https://doi.org/10.1145/2647868.2654970

[23] Yoshiyuki Kawano and Keiji Yanai. 2014. Food image recognition with deep convolutional features. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM, 589–593.

[24] 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, CEA@ACM Multimedia 2009, Beijing, China, October 23, 2009*. 23–30. https://doi.org/10.1145/1630995.1631001

[25] Alex Krizhevsky. 2014. One Weird Trick For Parallelizing Convolutional Neural Networks. *CoRR* abs/1404.5997 (2014). arXiv:1404.5997 http://arxiv.org/abs/1404.5997

[26] Min Lin, Qiang Chen, and Shuicheng Yan. 2014. Network In Network. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*. http://arxiv.org/abs/1312.4400

[27] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. 2014. Microsoft COCO: Common Objects in Context. In *Computer Vision - ECCV 2014 - 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V*. 740–755. https://doi.org/10.1007/978-3-319-10602-1_48

[28] Yuzhen Lu. 2016. Food Image Recognition by using Convolutional Neural Networks (CNNs). *arXiv preprint arXiv:1612.00983* (2016).

[29] Niki Martinel, Gian Luca Foresti, and Christian Micheloni. 2016. Wide-Slice Residual Networks for Food Recognition. *CoRR* abs/1612.06543 (2016). arXiv:1612.06543 http://arxiv.org/abs/1612.06543

[30] Yuji Matsuda, Hajime Hoashi, and Keiji Yanai. 2012. Recognition of Multiple-Food Images by Detecting Candidate Regions. In *Proceedings of the 2012 IEEE International Conference on Multimedia and Expo, ICME 2012, Melbourne, Australia, July 9-13, 2012*. 25–30. https://doi.org/10.1109/ICME.2012.157

[31] Yelena Mejova, Sofiane Abbar, and Hamed Haddadi. 2016. Fetishizing Food in Digital Age: #foodporn Around the World. In *Proceedings of the Tenth International Conference on Web and Social Media, Cologne, Germany, May 17-20, 2016*. 250–258. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/12998

[32] Yelena Mejova, Hamed Haddadi, Anastasios Noulas, and Ingmar Weber. 2015. # foodporn: Obesity Patterns in Culinary Interactions. In *Proceedings of the 5th international conference on digital health 2015*. ACM, 51–58.

[33] Michele Merler, Hui Wu, Rosario Uceda-Sosa, Quoc-Bao Nguyen, and John R. Smith. 2016. Snap, Eat, RepEat: A Food Recognition Engine for Dietary Logging. In *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management, MADiMa @ ACM Multimedia 2016, Amsterdam, The Netherlands, October 16, 2016*. 31–40. http://dl.acm.org/citation.cfm?id=2986036

[34] Weiqing Min, Bing-Kun Bao, Shuhuan Mei, Yaohui Zhu, Yong Rui, and Shuqiang Jiang. 2018. You Are What You Eat: Exploring Rich Recipe Information for Cross-Region Food Analysis. *IEEE Trans. Multimedia* 20, 4 (2018), 950–964. https://doi.org/10.1109/TMM.2017.2759499

[35] Ferda Ofli, Yusuf Aytar, Ingmar Weber, Raggi al Hammouri, and Antonio Torralba. 2017. Is Saki #delicious?: The Food Perception Gap on Instagram and Its Relation to Health. In *Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3-7, 2017*. 509–518. https://doi.org/10.1145/3038912.3052663

[36] World Health Organization. 2019. Healthy Diet, Last accessed on February, 2019. https://www.who.int/news-room/fact-sheets/detail/healthy-diet

[37] Thanh-Trung Phan and Daniel Gatica-Perez. 2017. Healthy #fondue #dinner: Analysis and Inference of Food and Drink Consumption Patterns on Instagram. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia, MUM 2017, Stuttgart, Germany, November 26 - 29, 2017*. 327–338. https://doi.org/10.1145/3152832.3152857

[38] Alison J Price, Amelia C Crampin, Alemayehu Amberbir, Ndoliwe Kayuni-Chihana, Crispin Musicha, Terence Tafatatha, Keith Branson, Debbie A Lawlor, Elenaus Mwaiyeghele, Lawrence Nkhwazi, et al. 2018. Prevalence of Obesity, Hypertension, and Diabetes, and Cascade of Care in Sub-Saharan Africa: A Cross-sectional, Population-based Study in Rural and Urban Malawi. *The lancet Diabetes & endocrinology* 6, 3 (2018), 208–222.

[39] Francesco Ragusa, Valeria Tomaselli, Antonino Furnari, Sebastiano Battiato, and Giovanni Maria Farinella. 2016. Food vs Non-Food Classification. In *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management,*

MADiMa @ ACM Multimedia 2016, Amsterdam, The Netherlands, October 16, 2016. 77–81. http://dl.acm.org/citation.cfm?id=2986041

[40] Joseph Redmon and Ali Farhadi. 2018. YOLOv3: An Incremental Improvement. *CoRR* abs/1804.02767 (2018). arXiv:1804.02767 http://arxiv.org/abs/1804.02767

[41] Jaclyn Rich, Hamed Haddadi, and Timothy M. Hospedales. 2016. Towards Bottom-Up Analysis of Social Food. In *Proceedings of the 6th International Conference on Digital Health Conference, DH 2016, Montréal, QC, Canada, April 11-13, 2016.* 111–120. https://doi.org/10.1145/2896338.2897734

[42] 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*. ACM, 115–116.

[43] 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, MADiMa @ ACM Multimedia 2016, Amsterdam, The Netherlands, October 16, 2016.* 3–11. http://dl.acm.org/citation.cfm?id=2986039

[44] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A. Alemi. 2017. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA.* 4278–4284. http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14806

[45] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott E. Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going Deeper with Convolutions. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015.* 1–9. https://doi.org/10.1109/CVPR.2015.7298594

[46] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. 2016. Rethinking the Inception Architecture for Computer Vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016.* 2818–2826. https://doi.org/10.1109/CVPR.2016.308

[47] Antonio Torralba and Alexei A. Efros. 2011. Unbiased Look at Dataset Bias. In *The 24th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, Colorado Springs, CO, USA, 20-25 June 2011.* 1521–1528. https://doi.org/10.1109/CVPR.2011.5995347

[48] Pixie G Turner and Carmen E Lefevre. 2017. Instagram use is Linked to Increased Symptoms of Orthorexia Nervosa. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity* 22, 2 (2017), 277–284.

[49] Xin Wang, Devinder Kumar, Nicolas Thome, Matthieu Cord, and Frédéric Precioso. 2015. Recipe Recognition with Large Multimodal Food Dataset. In *2015 IEEE International Conference on Multimedia & Expo Workshops, ICME Workshops 2015, Turin, Italy, June 29 - July 3, 2015.* 1–6. https://doi.org/10.1109/ICMEW.2015.7169757

[50] Saining Xie, Ross B. Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. 2017. Aggregated Residual Transformations for Deep Neural Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017.* 5987–5995. https://doi.org/10.1109/CVPR.2017.634

[51] Keiji Yanai and Yoshiyuki Kawano. 2015. Food Image Recognition using Deep Convolutional Network with Pre-training and Fine-tuning. In *2015 IEEE International Conference on Multimedia & Expo Workshops, ICME Workshops 2015, Turin, Italy, June 29 - July 3, 2015.* 1–6. https://doi.org/10.1109/ICMEW.2015.7169816

[52] Agneta Yngve, Barrie Margetts, Roger Hughes, and Marilyn Tseng. 2009. Food Insecurity–Not Just about Rural Communities in Africa and Asia. *Public health nutrition* 12, 11 (2009), 1971–1972.