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

ACM MM MADiMa Workshop 2019

5th International Workshop on Multimedia Assisted Dietary Management

Authors: Mona Jalal*, Kaihong Wang*, Sankara, Jefferson, Yi Zheng, Elaine Nsoesie, and Margrit Betke

* Represents co-first authors

Motivation

- Create a food/non-food dataset so that Kenyan food/non-food classifiers that can detect Kenyan food
- Create a food type dataset to recognize Kenyan food types
- Using the metadata in social Instagram post for creating multimodal classifiers
- Why is our food type detection method that also incorporate the information from the caption is interesting and superior to other methods?

BOSTON UNIVERSITY LORI

Introduction

- We propose Scrape-by-Location system and Scrape-by-Keywords system to collect two datasets: **Kenyan104K** and **KenyanFood13**.
- We propose a food/non-food classifiers trained on Kenyan104K: **KenyanFC** and a multimodal food type classifier trained on KenyanFood13: **KenyanFTR**.
- We apply our techniques to millions of Instagram posts and analyze food pattern in Kenya.

BOSTON UNIVERSITY

LORI

Related Work

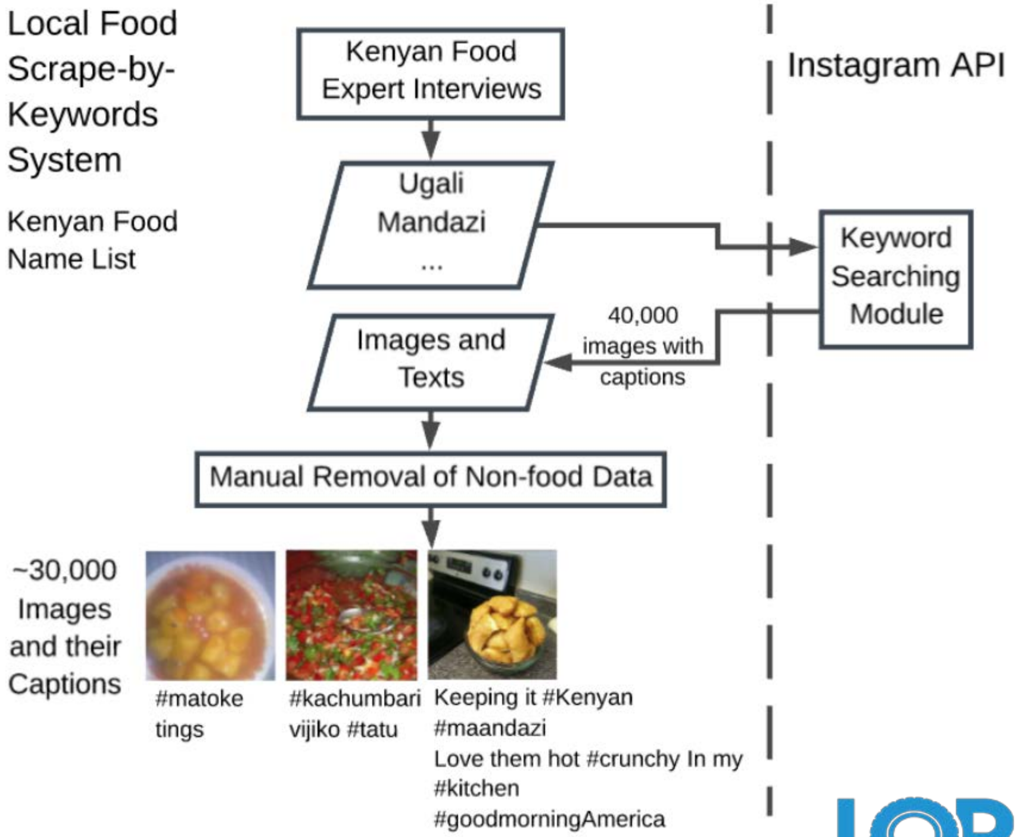
Food type datasets

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

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

Proposed Scrape-by-Keyword System

- 38 initial keywords
- 40,000 initial images
- Manually removed non-food images
- Finally ended up with 30K Kenyan foods
- Merge keywords with duplicate meaning
- For each post we downloaded the:
  - Image primary key,
  - Image(s) and Image URL(s)
  - Caption (if exists)



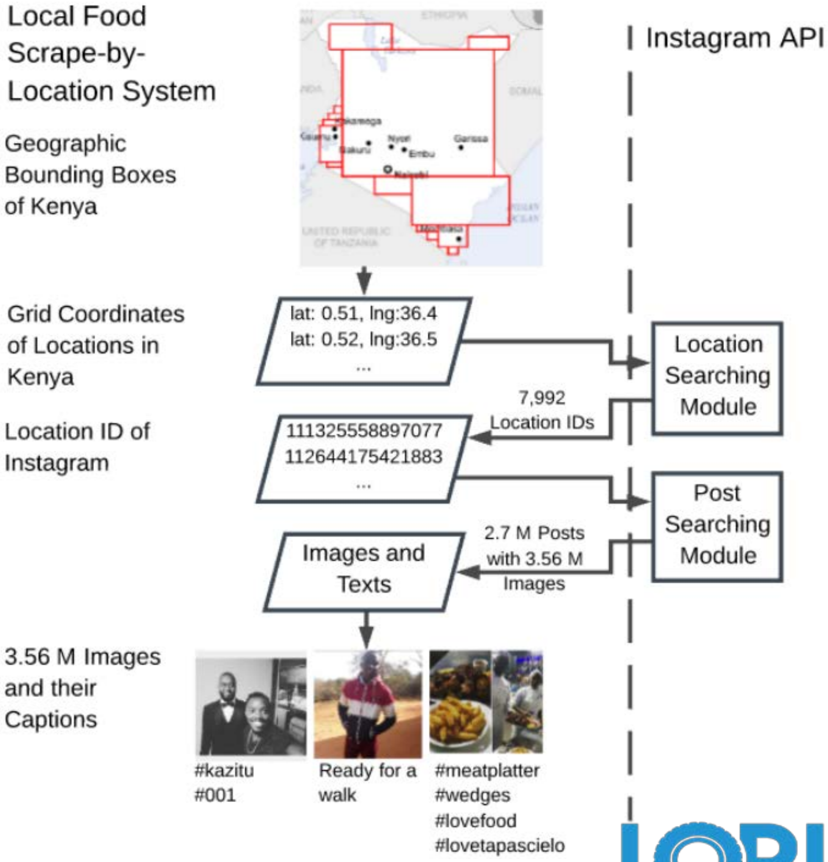
BOSTON UNIVERSITY

LORI

Proposed Scrape-by-Location System

A total of 3.56M images

- We defined a set of rectangular regions on Kenya
- For each point on the grid, we searched all nearby locations registered on Instagram for 20 days
- We retrieved the recent posts of each location ID
- For each post we downloaded the:
    - Image primary key,
    - Location ID
    - Image(s) and Image URL(s)
    - Caption (if exists)
    - Latitude and longitude of the location

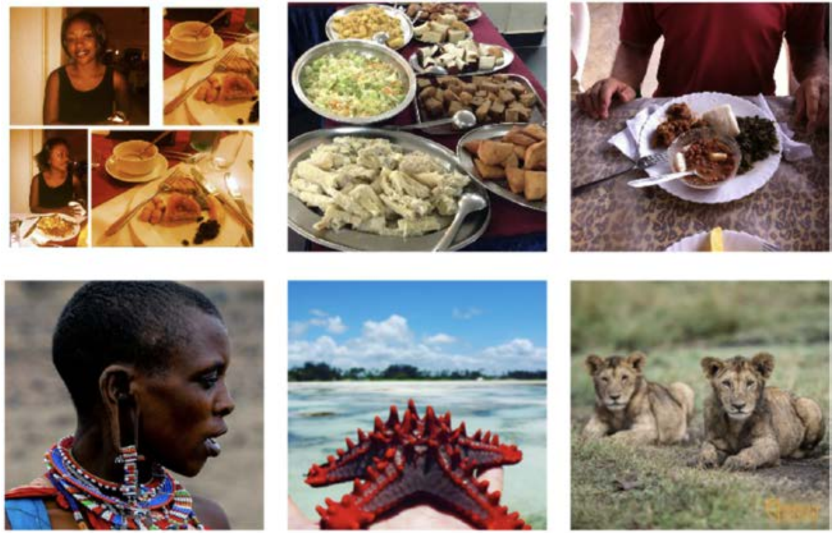


BOSTON UNIVERSITY

LORI

Kenyan Food/Non-Food Dataset (Kenya104K)

- Kenyan104K has 52K food and 52K non-food images
- Adapted for classifying Kenyan food/non-food images
- Consists of 30,000 non-food images collected by our Scrape-by-Location method and 30,000 food images by our Scrape-by-Keyword system, both with manual inspection, along with 9,658 food images from Instagram Food Dataset (IFD) and 53,527 food and non-food images from FCD

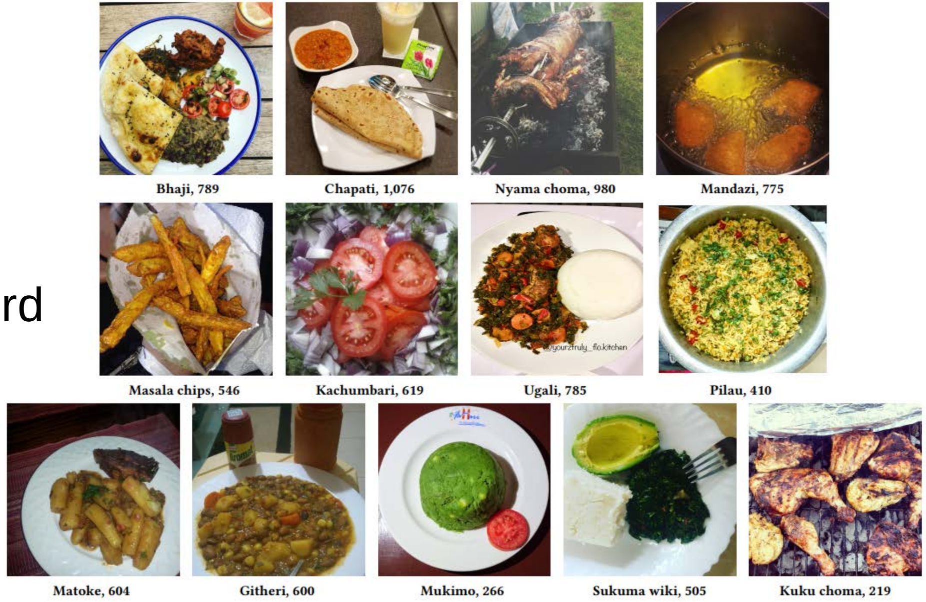


BOSTON UNIVERSITY

LORI

Kenyan Food Type Dataset

- 13 food types
- ~629 images per class
- 8,174 images in total
- Created from an initial pool of 30K images retrieved by Scrape-by-Keyword methodology where “keywords” were popular Kenyan foods.
- Only 15 of the initial 38 Kenyan foods had more than 500 images in Instagram.
- There is no guarantee that images of a hashtag would contain the specific food



→ We performed expert manual inspection for finalizing the food type dataset

BOSTON UNIVERSITY

LORI

Architecture of Food Type Recognition

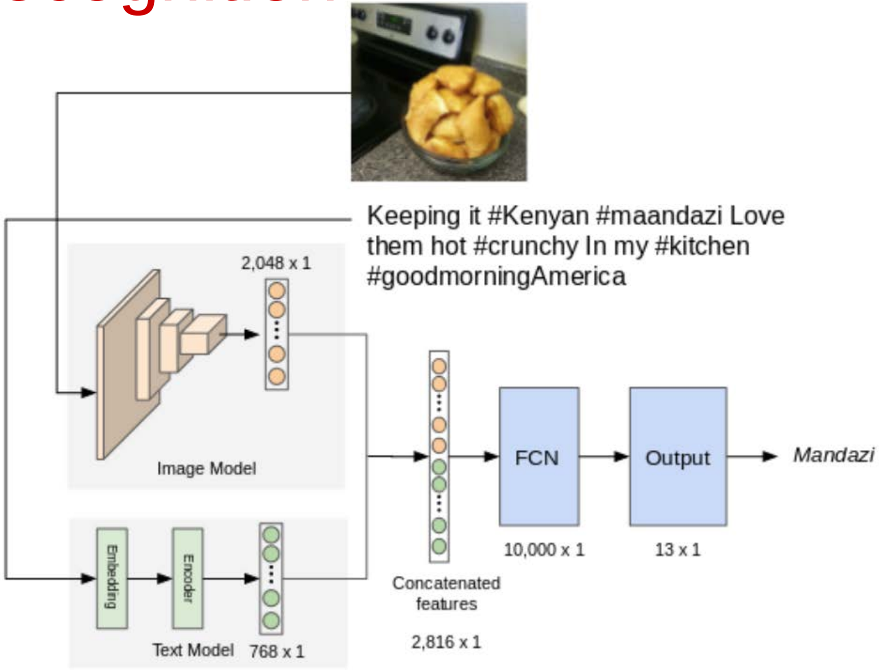
- We use BERT to extract feature vectors

from image captions (length: 768)

- Image feature vector is created by

ResNext101 (length: 2,048)

  - Pretrained on ImageNet dataset

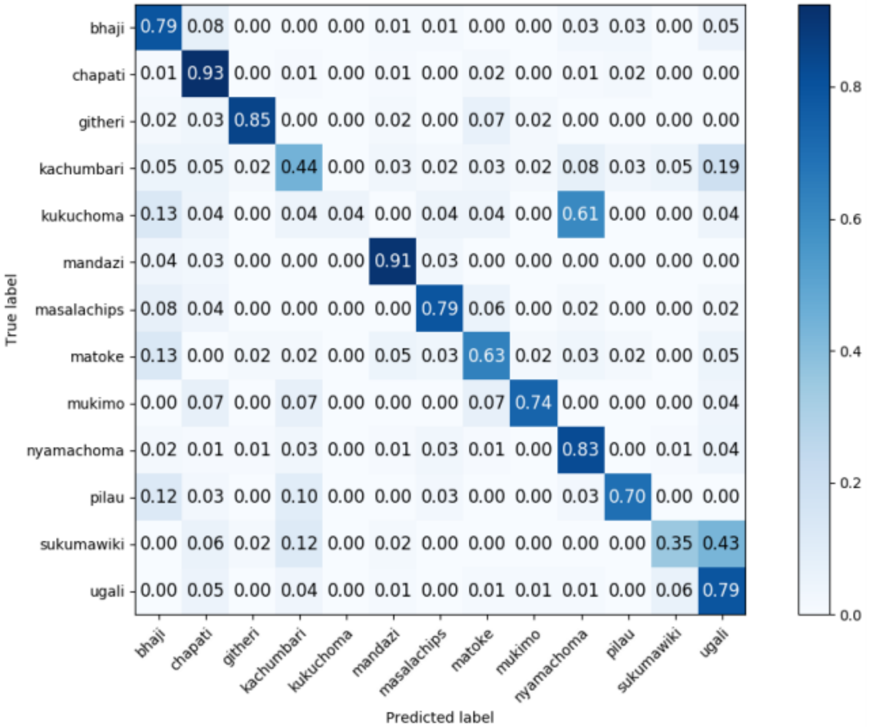


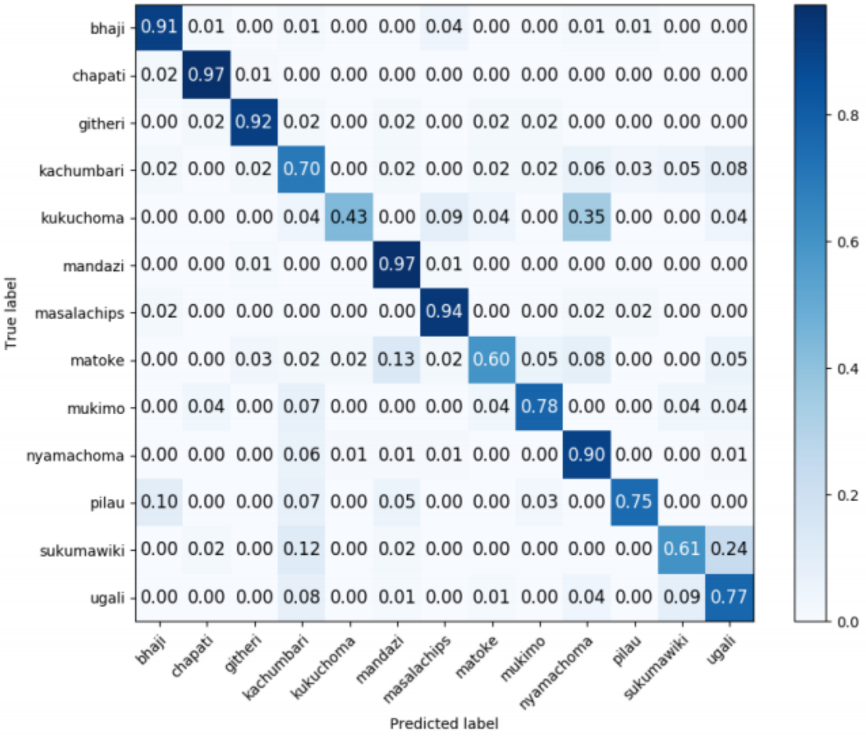
→ Eventually, we concatenate the language and image feature vectors using a fully connected network with 10K neurons and then pass it to an output layer with 13 classes.

BOSTON UNIVERSITY

LORI

Kenyan Food Type Recognition with/without Text





Some Preliminary Results 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%
KenyanFTR: Image + Caption	81.04%± 0.86%	95.95%± 0.44%

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%
KenyanFTR: ResNeXt101+BERT	81.04%± 0.86%	95.95%± 0.44%

BOSTON UNIVERSITY

Misclassifications when Trained on FCD and Tested on Kenyan104K

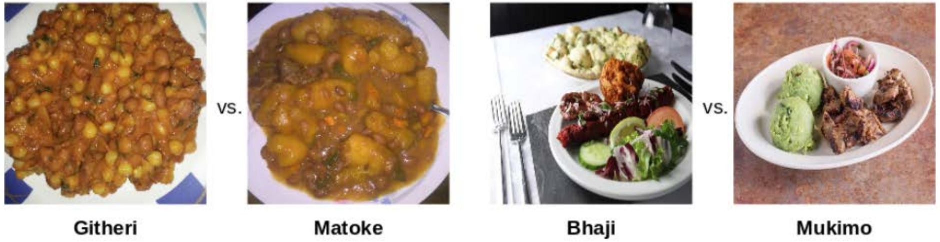
Non-food detected as food

Food detected as non-food

BOSTON UNIVERSITY

LORI

Most Confused Classes in KenyanFood13

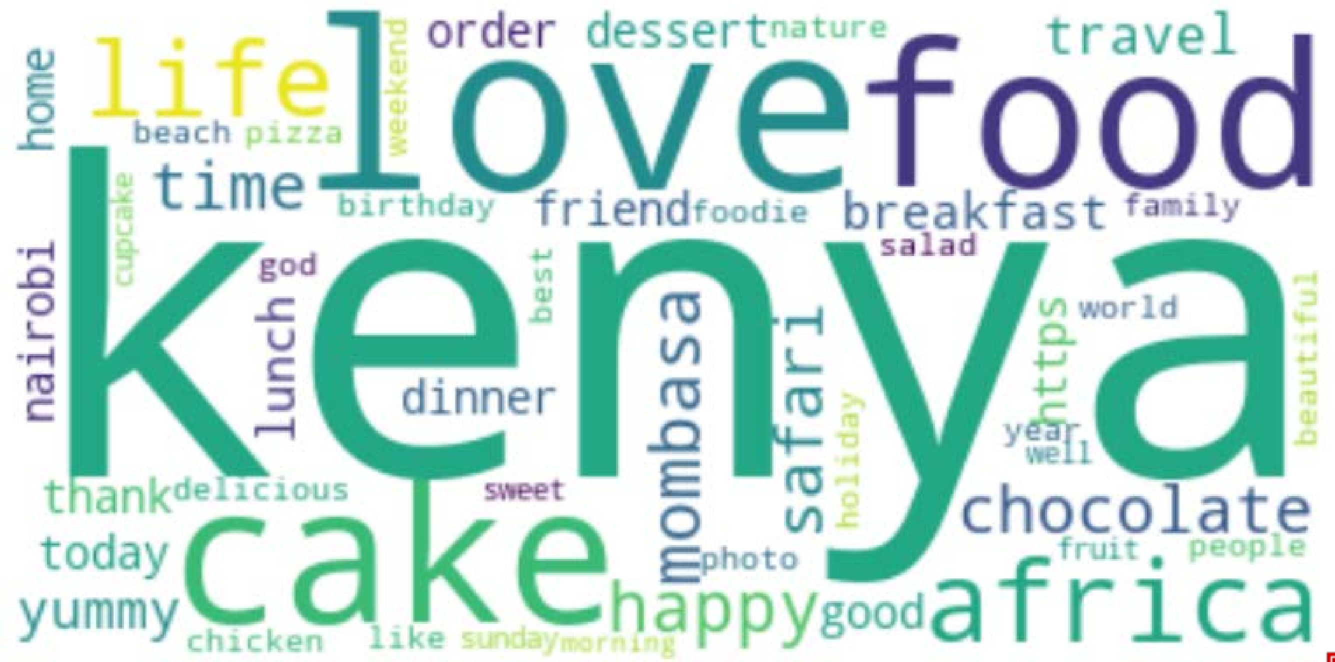


- Computed the feature vector by chopping the last layer of a ResNet50
- Computed the L2 distance between each pair of feature vectors
- For example, Githeri and Matoke are visually very similar hence the L2 distance of their feature vectors are the smallest

BOSTON UNIVERSITY

LORI

Thank You



BOSTON UNIVERSITY

LORI