Imperial College London

A computer vision-based passive dietary intake monitoring system for assessing dietary intake in low-and middle-income countries

Jianing Qiu, **Frank Lo**, and Benny Lo

Hamlyn Centre for Robotic Surgery, Imperial College London

Abstract:

Malnutrition is a major health challenge in many low- and middle-income countries (LMICs). Accurate dietary intake assessment is important to reduce malnutrition. However, existing dietary assessment tools require access to a computer/smartphone, limiting their deployment in resource-poor LMICs.

Method 1: Passive Dietary monitoring

- Using low-cost **wearable cameras** in measuring dietary intake through passive dietary monitoring method (i.e., AI-based).
- **Raw videos** are fed into our proposed model.
- Consumed **food recognition** and **bite counting** can be inferred to provide guidance to dietitians for dietary assessment.

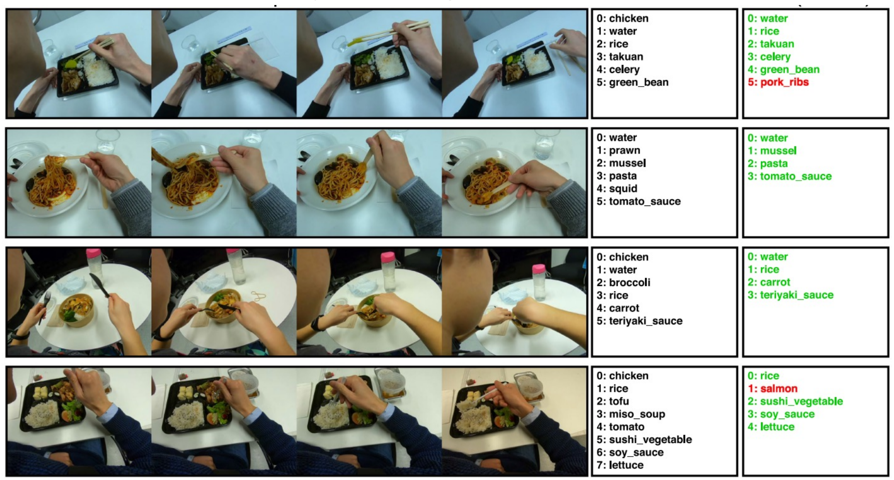

Fig 1. Recognized food items (ingredients and drinks)

potato	chicken	beef	soda	water	orange_juice
red_wine	champagne	green_onion	seaweed	dumpling	broccoli
napa_cabbage	chili_pepper	rice	chicken_katsu	curry	pickled_radish

Fig 2. 66 unique food items labelled in the dataset

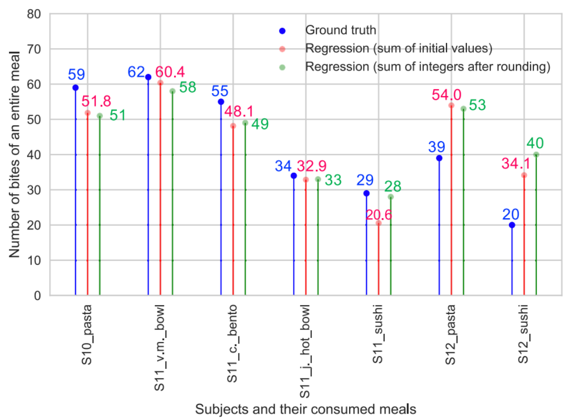

Fig 3. Qualitative results of bite counting

Results:

ACC in Food Recognition (Board Categories): 97.55%
ACC in Food Recognition (Fine Categories): 54.77
Error in bite count (Mean square error): 0.312

Qiu, Jianing, et al. "Counting bites and recognizing consumed food from videos for passive dietary monitoring." *IEEE Journal of Biomedical and Health Informatics* 25.5 (2020): 1471-1482.

Method 2: Image Captioning

The subject is eating rice with egg stew with the family, and the bowl is half empty

1. Food categories
2. Eating alone/with others
3. A rough portion size

- A novel **captioning model** is also designed to generate the captions for the dietary images.
- Advantages:
  1. Preserve the subject's privacy
  2. Reduce workload (i.e., frames with eating episodes can be obtained easily)
  3. Can generate a nutrient intake report with consumed food type, eating duration, and rough volume.

Examples

Up-Down: the subject is eating jollof rice
Att2in: the subject is eating jollof rice
$\mathcal{M}^2$ **Transformer**: the subject is eating jollof rice
GL-Transformer*: the subject is eating jollof rice , and more than half of the bowl is empty
GL-Transformer: the subject is eating jollof rice , and only less than half is left in the bowl
GT: the subject is eating jollof rice, and there isn't much left in the bowl

Up-Down: the subject is processing fish in the kitchen
Att2in: the subject is having a meal
$\mathcal{M}^2$ **Transformer**: the subject is eating akple and okra stew
GL-Transformer*: the subject is eating akple and okra stew with the children
GL-Transformer: the subject is processing some green vegetables
GT: the subject is processing some green vegetables

Qiu, Jianing, Frank P-W. Lo, Xiao Gu, Modou L. Jobarteh, Wenyan Jia, Tom Baranowski, Matilda Steiner-Asiedu et al. "Egocentric Image Captioning for Privacy-Preserved Passive Dietary Intake Monitoring." *arXiv preprint arXiv:2107.00372* (2021).