This paper presents the design and development of an integrated system for mobile image-based dietary assessment. The system is designed to capture and analyze food images taken by study participants in naturalistic settings. The main components of the system include a mobile Food Record (mFR), a secure cloud-based server, and a web interface. The mFR captures images of food items before and after eating, along with metadata such as date, time, GPS coordinates, and camera pose and angle. These images are sent to the cloud server, where they are segmented, and food types are classified using image analysis. The system is supported by an integrated database system, which allows nutrition researchers to visualize and verify collected food images and the associated dietary information. The contributions of this paper are twofold. First, we propose a system that provides an efficient mechanism for image-based dietary assessment to support nutrition science researchers. Second, the proposed system is capable of building a comprehensive food image dataset with high quality visual and nutrition groundtruth for developing novel computational approaches.
interface for the participant to review, modify and confirm the food labels received from the server (Step 4). Participant confirmed results are then sent back to the server for additional refinement and portion size estimation (Step 5). Results can then be viewed and interacted with via the web interface by nutrition researchers from any computer with an internet connection to perform nutrition analysis (Step 6).

2.1 Mobile Food Record
The Mobile Food Record (mFR) allows participant to record the eating occasion images and review the image analysis results from the server. Participants record the foods and beverages consumed at a meal by capturing a pair of images, one image before eating and one image after eating. This pair of images are sent to the server for analysis. Results are sent back to the mFR for participants to review, modify, and confirm. This information can be used by the image analysis done on the server to update the learning methods. The review process requires participant to confirm or modify the food labels by clicking the pins on the before image (Figure 2a). A list of foods relevant to the nutrition study is pre-loaded on the mFR to allow participants to choose from (Figure 2b). Details on the mFR interface design and functionalities can be found in [5].

2.2 Cloud Server
The cloud server is responsible for the data transmission, storage and inference. The server receives the eating occasion images and associated metadata from the mFR. This information is stored in an integrated database system [7]. The cloud server provides preliminary image analysis results that are sent back to the mFR for the participant to review and confirm. Results are further refined, from which portion estimation and nutrition analysis are performed. Image analysis and inference include food segmentation, food classification and portion size estimation [14, 23, 39]. This information can be visualized and used for additional nutrition analysis using the web interface by nutrition researchers.

2.3 Web Interface
In addition to visualizing the collected data from dietary studies, we designed a set of new features for the web interface and incorporated them into the TADA system. Our goal is to design web-based annotation tools that can be interacted by nutrition researchers to review and modify participant confirmed results and to conduct additional nutrition analysis. The annotation tool provides a list
of foods with associated food codes from the USDA Food and Nutrition Database for Dietary Studies (FNDDS) database [3] and a searching tool to quickly return relevant food labels. Our design takes into consideration the following criteria including easy and quick accessibility for researchers to modify the preliminary results from the participants, and high quality annotation results that can be used to improve the accuracy of the image analysis.

The web-based annotation tools consist of two key "steps" for researchers to access and modify participant confirmed preliminary results. These steps include review of collected food image and its associated dietary information, and the modification of participant confirmed preliminary results.

Review of collected food image and associated dietary information. For each participant, all captured before and after eating scene images are displayed as small preview images on the web interface. Researchers can click either the before or after eating scene preview image to review images captured for that eating occasion, along with participant confirmed food types and researcher confirmed final annotations as shown in Figure 3. The annotations from participants and researchers are displayed on the eating scene image with the food labels and bounding boxes information. An example is shown in Figure 3 where some foods are not labeled correctly by the participants and the bounding box is not accurately. The web interface provides an interactive design to modify the participant confirmed preliminary results.

Modification of participant confirmed preliminary results. The web-based annotation tool has a dedicated view for the researcher to modify the annotations received from participants. This function can be accessed by clicking the "edit" button below the researcher annotation as shown Figure 3). The researcher annotation process consists of three main operations: entering the initials of the researcher performing the task, drawing a bounding box around the food item in the image, selecting the corresponding food label to save the bounding box. The process can be repeated if there are multiple foods in the eating scene image.

As shown in Figure 4a, researchers can modify or add a food by drawing a bounding box on the eating scene image using the hold-and-drag operation common to all operating systems. The instruction for this step is displayed on the top of the web interface to provide clear guideline to the researchers. Once the bounding box is drawn, a cropped image of the extract region is shown in the review section. This allows researchers to verify the bounding box is accurately draw, particularly for more complex eating scenes with multiple food items.

The researcher can select a food label to match the cropped region of the image by using the search bar on the web interface, which contains a pre-loaded food list and associated USDA FNDDS food codes. The search can be done by typing in either the food name or the food code in the search bar. The search mechanism is similar to what is used in most operating systems which shows all relevant results containing the keywords. For example, when 'potato' is typed, then 'potato', 'potato wedges' and 'roast potato' are shown in search bar. Because some food names are the same, the matched food codes for the food names are shown to differentiate them. If the food label is not on the pre-loaded food list, the researcher can click the save button directly and the keyword entered will be saved as the food label for the bounding box. Once the food label is selected for the bounding box, corresponding energy information can be entered manually by the researcher. An example of the saved food region, matched food name and nutrition information are shown in Figure 4b). A delete option is provided if there is any mistake made during this process. Once the research can click the Finish button at the bottom of the web page once information for all foods in an image is completed. The aggregated information from the researcher and the participant is made available for download.

3 DEPLOYMENT IN DIETARY STUDIES

One main goal of our system is to provide nutrition researchers an efficient mechanism to collect high quality eating occasion images and nutrition information for dietary assessment. The system has been widely deployed in more than 30 dietary studies with over 2,500 participants between the ages 6 months – 70 years in domestic and international locations. We have collected more than 72,000 images from both controlled feeding and community dwelling studies. Table 1 describes some of the published studies. Our system has shown to be acceptable and feasible for dietary studies as indicated in [9] where participants were asked to capture images of all eating occasions over 7.5 days. Prior to using the mFR, 71% of them agreed “Remembering to take an image before meals would be easy.” After using the mFR for 7.5 days, the agreement rate reached 100%. For capturing food images, participants were provided with a known-dimension fiducial marker placed in eating scene for providing
We observed that estimating food energy accurately from an eating scene image is a challenging task for people who do not have domain knowledge, note that some people were not able to recall the eating event. We have developed vision-based solutions to many challenging aspects of novel image-based dietary assessment tools. On food recognition, we developed a two-step approach for food localization and hierarchical food classification using Convolutional Neural Networks (CNNs) to reduce prediction error for visually similar foods [23], and continual learning in the challenging online learning scenario that is further bounded by run-time and limited data [18, 20]. On image segmentation, we developed class-agnostic method using a pair of eating scene images to find the salient missing objects without prior information about the food class [41] and developed weakly supervised and efficient superpixel based methods [38, 39]. On food portion size estimation, we developed different approaches using a single-view image, including the use of geometric models to recover 3D parameters of food objects in the scene [12], incorporating co-occurrence patterns to refine portion estimation results [15], and deep learning for the mapping of food images to food energy [13, 14]. To understand the contextual and dynamic attributes of diet, we explored various aspects of eating context and environment that influence dietary intake, such as combining temporal information and food co-occurrence to develop personalized learning model [37], and understanding different eating environments through novel image clustering [40].

In addition to developing methods for individual image analysis tasks, We have also developed integrated food image analysis systems including multiple hypothesis approach [42], and deep learning based methods that jointly performs food localization, classification, and portion size estimation [17, 19]. Figure 5 shows results of improved portion estimation from different approaches. We observed that estimating food energy accurately from an eating scene image is a challenging task for people who do not have domain knowledge, note that some people were not able to recall items they consumed. Trained on data collected by our proposed

### Table 1: Summary of published studies using the system.

| Study | Location | Population | Age          |
|-------|----------|------------|--------------|
| [4]   | Guam     | 65         | 3-10 years   |
| [6]   | Perth, West Australia | 58 | 12-30 years |
| [9]   | Tippecanoe County, IN | 45 | 21-65 years |
| [8]   | Pacific Coast, Washington | 24 | 21-60 years |
| [10]  | Tippecanoe County, IN | 57 | 21-65 years |
| [21]  | Perth, West Australia | 247 | 18-30 years |
| [16]  | Perth, West Australia | 165 | 18-65 years |
| [25]  | O’ahu     | 93         | 9-13 years   |
| [26]  | O’ahu     | 60         | 35-55 years  |
| [27]  | Milwaukee, WI | 12 | 8-18 years |
| [34]  | Tippecanoe County, IN | 63 | 11-18 years |
| [35]  | Perth, West Australia | 118 | 25-70 years |

### 4 FOOD IMAGE ANALYSIS

Accurate estimation of dietary intake relies on the system’s ability to distinguish foods from image background (i.e., localization and segmentation), to identify or label the foods (i.e., classification), to estimate food portion size, and to understand the context of the eating event. We have developed vision-based solutions to many challenging aspects of novel image-based dietary assessment tools. On food recognition, we developed a two-step approach for food localization and hierarchical food classification using Convolutional Neural Networks (CNNs) to reduce prediction error for visually similar foods [23], and continual learning in the challenging online learning scenario that is further bounded by run-time and limited data [18, 20]. On image segmentation, we developed class-agnostic method using a pair of eating scene images to find the salient missing objects without prior information about the food class [41] and developed weakly supervised and efficient superpixel based methods [38, 39]. On food portion size estimation, we developed different approaches using a single-view image, including the use of geometric models to recover 3D parameters of food objects in the scene [12], incorporating co-occurrence patterns to refine portion estimation results [15], and deep learning for the mapping of food images to food energy [13, 14]. To understand the contextual and dynamic attributes of diet, we explored various aspects of eating context and environment that influence dietary intake, such as combining temporal information and food co-occurrence to develop personalized learning model [37], and understanding different eating environments through novel image clustering [40].

In addition to developing methods for individual image analysis tasks, we have also developed integrated food image analysis systems including multiple hypothesis approach [42], and deep learning based methods that jointly performs food localization, classification, and portion size estimation [17, 19]. Figure 5 shows results of improved portion estimation from different approaches. We observed that estimating food energy accurately from an eating scene image is a challenging task for people who do not have domain knowledge, note that some people were not able to recall items they consumed. Trained on data collected by our proposed

### 5 CONCLUSION

We enhanced a previously proposed image-based dietary assessment system by implementing new functionalities to efficiently collect food images and generate high quality visual and nutrition groundtruth information in a systemic design. Our system provides the necessary tools for researchers to conduct both image analysis and nutrition research more efficiently. We envision this system would open doors to many possible image-based dietary assessment applications using deep learning techniques. Our integrated system holds promise to improving the system efficiency and accuracy, reducing both participant and researcher burden associated with traditional approaches.

### REFERENCES

[1] [n.d.]. Amazon Mechanical Turk. https://www.mturk.com/.
[2] [n.d.]. CDC Cost. https://www.cdc.gov/chronicdisease/about/costs/index.htm. Jan, 2021.
[3] [n.d.]. USDA Food and Nutrient Database for Dietary Studies 2017-2018. Food Surveys Research Group Home Page, http://www.ars.usda.gov/nea/bhnrc/fsrg. Jan, 2021.
[4] T. F. Aflague, C. J. Boushey, R. T. L. Guerrero, Z. Ahmad, D. A. Kerr, and E. J. Delp. 2015. Feasibility and use of the mobile food record for capturing eating occasions among children ages 3–10 years in Guam. *Nutrients* 7, 6 (2015), 4401–4415.
[5] Z. Ahmad, M. Bosch, N. Khanna, D. Kerr, C. Boushey, F. Zhu, and E. Delp. 2016. A Mobile Food Record For Integrated Dietary Assessment. *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management* (October 2016), 53–62. Amsterdam, Netherlands.
[6] K. E. Bathgate, J. L. Sherriff, H. Leonard, S. S. Dhaliwal, E. J. Delp, C. J. Boushey, and D. A. Kerr. 2017. Feasibility of assessing diet with a mobile food record for adolescents and young adults with down syndrome. *Nutrients* 9, 3 (2017), 273.
[7] Marc Bosch, TusaRebecca Schap, Fengqing Zhu, Nitin Khanna, Carol J. Boushey, and Edward J. Delp. 2011. Integrated database system for mobile dietary assessment and analysis. *Proceedings of IEEE International Conference on Multimedia and Expo* (Jul 2011), 1–6.
