AIFood: A Large Scale Food Images Dataset for Ingredient Recognition

Gwo Giun (Chris) Lee*, Chin-Wei Huang*, Jia-Hong Chen*, Shih-Yu Chen*, Hsiu-Ling Chen*
*National Cheng Kung University, Tainan, Taiwan
E-mail: clee@mail.ncku.edu.tw Tel: +886-6-2757575 ext.62448

***Abstract*— In this paper, we introduce a large-scale food images dataset namely AIFood, which is constructed to aim ingredient recognition in food image research. AIFood dataset includes 24 categories and totally 372,095 food images around the world. We collect food images from eight existing food image datasets and a food website. The food images are relabeled using 24 categories. We preliminarily label each image using existing food information, e.g. dish name or ingredient information. Next, we manually check food images to find out undiscovered ingredients and relabel them. Every image can be labeled more than one category. In addition, food images may have color cast or uneven contrast problems, which may disturb performance of image recognition system. So, we applied preprocessing method which contains automatic white balancing and contrast limited adaptive histogram equalization methods to improve visual quality of food images. We set constraints which are defined by luminance and chrominance of image to determine if the image is to be preprocessed.**

I. INTRODUCTION

Food plays an important role in people's daily life. The kinds of food are different according to the culture and environment of the nationalities. Nowadays, people pay more and more attention to health. Dietary assessment is one of important topic of health. Dietary assessment can estimate nutrient and foods people intake every day so that people can understand what nutrients are lacking or excessive. Dietary assessment system can be constructed on cloud or edge device. For the convenience and personal privacy, dietary assessment system is usually applied to edge device like mobile phone. People just need to take the food they eat to food image recognition system to detect images and estimate the nutrient then perform dietary assessment and management on edge device. In order to have a good dietary assessment system, food recognition and nutrition estimation research have been gradually developed in recent years. There have several methods of feature extraction and classification or deep learning methods have been proposed. A strong food recognition system not only needs an available algorithm but also needs to have a large-scale food image dataset and corresponding ground truth label. There have some papers discuss and summarize methods and datasets of food image recognition [17] [18].

However, a large-scale food images and their accurate label are hard to acquire. Fortunately, since 2009, many food image datasets have been published and we can obtain those datasets from their own home pages or sending request email to authors. Those food images datasets are great contribution to food recognition research.

Many existing food image datasets are categorized according to dish name, that is, only one dish name label in one image. However, the number of dish is more than we think so that we can't collect every dish in the world. Moreover, every image in same dish may be composed of different ingredient and seasoning. We think that ingredient information in food images can help researchers to estimate nutrition accurately than only dish name.

In this paper, we introduce a food image dataset, namely AIFood, with ingredient label for ingredient recognition and nutrition estimation research. AIFood dataset contains 372,095 images and each image is multi-labeling and categorized to 24 categories. The 24 categories are food ingredient and can almost cover all kinds of people's daily diet. In addition, food images are usually taken by mobile phone or camera in various environments. Many food images may have color cast or uneven contrast problems due to environment or camera parameters. Those problems may cause food image recognition system to make a wrong prediction. Therefore, we preprocessed food images to handle two problem based on some constraints.

In the following sections, we introduce existing food image datasets. Next, we describe our labeling method and how we collected our food images of dataset. Finally, we introduce our preprocessing steps.

II. RELATED WORK

Dietary assessment and food research have been gradually developed since approximately 2009. There have several existing food image datasets have been proposed since 2009.

In 2009, Pittsburgh fast-food image dataset (PFID) has been proposed by Mei Chen et al [1]. PFID contains 4,545

978-1-7281-1895-6/19/$31.00 ©2019 IEEE

Authorized licensed use limited to: Malta College of ArtsScience&Tech. Downloaded on May 27,2021 at 11:50:24 UTC from IEEE Xplore. Restrictions apply.

images of 101 fast foods from 11 famous fast food restaurants including hamburgers, pizza, salad, etc.

In 2010, Hajime Hoashi et al. proposed Food85 dataset which contains 8,500 images of 85 common dishes in Japan [2].

In 2012, Yuji Matsuda et al. proposed the UECFood100 dataset which contains 9,060 images of 100 Japanese dishes [3]. They proposed the UECFood256 dataset based on the UEC Food100 in 2014. UECFood256 contains 31,397 images of 256 categories [4].

In 2014, Farinella et al. proposed UNICT-FD889 dataset composed by 3,583 images related to 889 dishes from different nationalities [5]. Lukas Bossard et al. proposed Food-101 dataset which contains top 101 food categories from Foodspotting, which is a famous food website. Each food category consists of 1,000 images, that is 101,000 images totally [6].

In 2015, Xin Wang et al. proposed the UPMC Food-101 which has same food categories as Food-101 and collected 90,840 images from Google Image search [7]. Oscar Beijbom et al. proposed Menu-match dataset which contains a total of 646 images tagged by 41 categories. The images and food information are acquired from Asian, Italian and soup restaurants [8]. Parisa Pouladzadeh et al. proposed FooDD dataset which contains around 3.000 images of 20 categories. FooDD dataset offers variety of food photos taken from different camera with different illuminations and multiple shooting angles [9]. The UNIMIB2015 and the UNIMINB2016 datasets are proposed by Ciocca et al. in 2015 and 2016 respectively. The UNIMIN2015 dataset is composed of 2,000 tray images with multiple foods and contains 15 food categories. On the other hand, the UNIMIB2016 dataset is composed of 1,027 tray images with multiple foods and contains 73 food categories. The images were acquired from canteen situated within the University of MilanoBicocca Campus [10].

In 2016, Ashutosh Singla et al. proposed Food5k and Food11 to perform food and non-food image classification and food categorization respectively [11]. Food5k contains 2500 food images and 2500 non-food images. Food11 contains 16,643 images of 11 categories. Jingjing Chen proposed Vireo Food-172 which contains 110,241 images with 172 dish categories and 353 ingredients. Each image has one dish name and complete ingredients in the image [12].

In 2017, Yanchao Liang et al. proposed the ECUSTFD dataset which contains 19 kinds of food and 2,978 images with a coin as calibration and top view and side view of food to estimation food volume and calorie [13]. Xin Chen proposed ChineseFoodNet dataset which contains 185,628 images of 208 Chinese dishes [14]. Jun Harashima et al. proposed Cookpad which contains about 1.6 million images after cooking. Cookpad also contains information about recipe, cooking process and process images [15]. Javier Marin proposed Recipe1M which contains over 1 million recipes and 80K images [16].

These datasets provide large-scale images and some of them provide recipe and ingredient information for food research. However, most of datasets don't provide complete ingredient information. So, we introduce a labeling method which can combine all datasets and provide ingredient information for ingredient recognition research.

Fig. 1 Example images from AIFood dataset. Every images have ingredient labels and have been preprocessed.

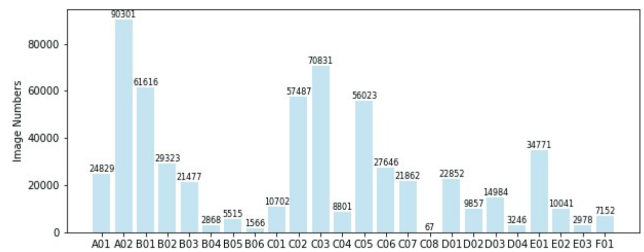

Fig. 2 Image number of 24 categories.

III. AIFOOD IMAGE DATASET

In this paper, we introduce the AIFood images dataset which contains 24 categories and 372,095 food images around the world. Instead of category of dish name, images of AIFood are labeled ingredient information. We describe how we formulate labeling method and collect food images as follows.

A. *Labeling Method*

Totally, we define 24 categories to represent ingredients in the food images, which basically cover the kinds of people's daily diet. Most of dishes can be represented by the 24 categories. For convenience, we also code 24 categories using an uppercase letter and number. The following is code name and ingredient name of the 24 categories: *(1) A01, Rice, (2) A02, Flour products, (3) B01, Pork, (4) B02, Chicken, (5) B03, Beef, (6), B04, Mutton, (7) B05, Duck, (8) B06, Goose, (9) C01, Fruit, (10) C02, Leafy vegetable, (11) C03, Pepo, (12) C04, Flower vegetable/Petals, (13) C05, Rhizome, (14) C06, Bean/Nut, (15) C07, Seaweed/Mushroom, (16) C08, Konjac, (17) D01, Fish, (18) D02, Shellfish, (19) D03, Crustacean, (20) D04, Cephalopod/Mollusk, (21) E01, Egg, (22) E02, Tofu, (23) E03, Hard bean curd/Soy milk film, (24) F01, Dairy products.*

Authorized licensed use limited to: Malta College of ArtsScience&Tech. Downloaded on May 27,2021 at 11:50:24 UTC from IEEE Xplore. Restrictions apply.

Example images with ingredient label from AIFood dataset are shown in Fig. 1. The image number of each category is shown in Fig. 2.

B. Data Collection

The food images of dataset were collected from existing food image datasets including ChineseFoodNet [14], Food-101 [6], ECUSTFD [13], Menu-match [8], UNIMIB2016 [10], FooDD [9], Food-11 [11], Vireo172 [12]. In addition, we downloaded food images from a cooking website ytower[1] to expand our dataset. Next, we labeled every food image using 24 categories.

Except for Vireo172 dataset which contains complete ingredient information of every image, the remaining datasets only have dish name rather than complete ingredient information in every image. For the images of Vireo172, we can simply transform the ingredient information into 24 categories and relabel the food images. For the remaining datasets, first, we make a preliminary ingredient mark for each dish name and label all images of each dish. For example, images labeling beef noodle are relabeled to A02 and B03. In this way, each image has at least one categories label. However, in some images, complete ingredients can't be labeled according to dish name. E.g. images of beef noodle may contain some vegetable or egg so that we can't label the complete ingredient according to dish name only. So, after the preliminary ingredient label according to dish name, manual confirmation is needed to correct those undiscovered ingredients in the food images. So far, all images of dataset have been labeled preliminary ingredient information and tens of thousands of images have been manually confirmed.

IV. PREPROCESSING

The food images are usually taken by mobile phone or camera in various environments. Color temperature and brightness in various environments affect visual appearance of food. Some of images may have serious color cast or uneven contrast problems. So, we preprocess the food image of AIFood dataset to solve those problems.

A. Automatic White Balancing

The color cast problem in food images may disturb food image recognition system. So, we applied an Automatic White Balancing (AWB) algorithm [19] to solve color cast problem. The AWB algorithm determines constraints on luminance and chrominance which are defined by the statistics from the HSI color space and the normalized RGB color space. the pixels in the image that conform to the constraints are considered as the potential white color. The potential white color areas in the image as reference to calculate the gain coefficient in the RGB color channels which can used to correct the colors to solve color cast problem in the image. In our dataset, we calculate the standard deviation of hue of potential white area in HSI space and average distance from white point in XYZ space to determine if this image should be corrected by AWB according to a threshold we set.

[1] https://www.ytower.com.tw/

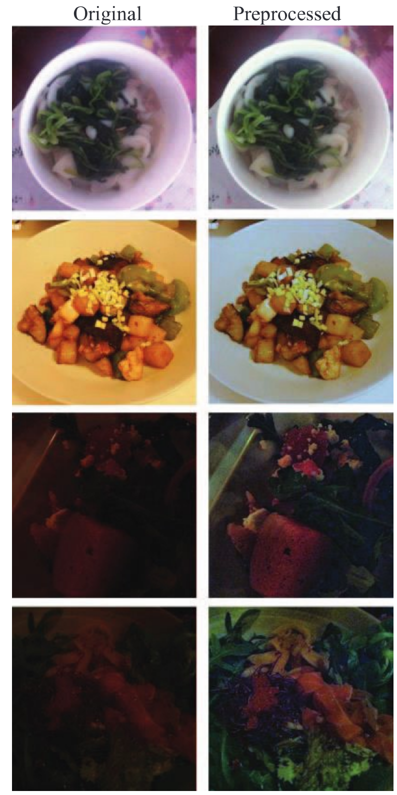

Fig. 3 Original and preprocessed images. We show four example images on each row. Left column are non-preprocessed images and right column are preprocessed images

B. Contrast Limited Adaptive Histogram Equalization

We applied the Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm to handle the too dark and too bright images. The CLAHE method uses a threshold to clip the histogram of intensity of an image and redistributes the histogram to improve the contrast in image [20]. In our dataset, we calculate the mean and standard deviation of intensity of an image and set threshold to determine if this image should be corrected by CLAHE method. In the implementation, we applied AWB first then CLAHE.

The original and preprocessed images are shown in Fig. 3. We can observe that the color cast problem have been eliminated and the preprocessed images are brighter and more vivid than original images on row 3 and row 4.

V. CONCLUSIONS

In this paper, we introduce a large-scale food image dataset for ingredient recognition which contains 372,095 food images. Each image has been labeled using 24 categories. In addition, we use AWB and CLAHE methods to improve

Authorized licensed use limited to: Malta College of ArtsScience&Tech. Downloaded on May 27,2021 at 11:50:24 UTC from IEEE Xplore. Restrictions apply.

visual quality of images. In future work, we will continue to relabel manually the images of dataset and expand our dataset.

References

[1] Chen, Mei, Kapil Dhingra, Wen Wu, Lei Yang, Rahul Sukthankar, and Jie Yang. "PFID: Pittsburgh fast-food image dataset." In 2009 16th IEEE International Conference on Image Processing (ICIP), pp. 289-292. IEEE, 2009.

[2] Hoashi, Hajime, Taichi Joutou, and Keiji Yanai. "Image recognition of 85 food categories by feature fusion." In 2010 IEEE International Symposium on Multimedia, pp. 296-301. IEEE, 2010.

[3] Matsuda, Yuji, Hajime Hoashi, and Keiji Yanai. "Recognition of multiple-food images by detecting candidate regions." In 2012 IEEE International Conference on Multimedia and Expo, pp. 25-30. IEEE, 2012.

[4] Kawano, Yoshiyuki, and Keiji Yanai. "Automatic expansion of a food image dataset leveraging existing categories with domain adaptation." In European Conference on Computer Vision, pp. 3-17. Springer, Cham, 2014.

[5] Farinella, Giovanni Maria, Dario Allegra, and Filippo Stanco. "A benchmark dataset to study the representation of food images." In European Conference on Computer Vision, pp. 584-599. Springer, Cham, 2014.

[6] Bossard, Lukas, Matthieu Guillaumin, and Luc Van Gool. "Food-101–mining discriminative components with random forests." In European Conference on Computer Vision, pp. 446-461. Springer, Cham, 2014.

[7] Wang, Xin, Devinder Kumar, Nicolas Thome, Matthieu Cord, and Frederic Precioso. "Recipe recognition with large multimodal food dataset." In 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), pp. 1-6. IEEE, 2015.

[8] Beijbom, Oscar, Neel Joshi, Dan Morris, Scott Saponas, and Siddharth Khullar. "Menu-match: Restaurant-specific food logging from images." In 2015 IEEE Winter Conference on Applications of Computer Vision, pp. 844-851. IEEE, 2015.

[9] Pouladzadeh, Parisa, Abdulsalam Yassine, and Shervin Shirmohammadi. "Foodd: food detection dataset for calorie measurement using food images." In International Conference on Image Analysis and Processing, pp. 441-448. Springer, Cham, 2015.

[10] Ciocca, Gianluigi, Paolo Napoletano, and Raimondo Schettini. "Food recognition: a new dataset, experiments, and results." IEEE journal of biomedical and health informatics 21, no. 3 (2016): 588-598.

[11] Singla, Ashutosh, Lin Yuan, and Touradj Ebrahimi. "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, pp. 3-11. ACM, 2016.

[12] Chen, Jingjing, and Chong-Wah Ngo. "Deep-based ingredient recognition for cooking recipe retrieval." In Proceedings of the 24th ACM international conference on Multimedia, pp. 32-41. ACM, 2016.

[13] Liang, Yanchao, and Jianhua Li. "Computer vision-based food calorie estimation: dataset, method, and experiment." arXiv preprint arXiv:1705.07632 (2017).

[14] Chen, Xin, Yu Zhu, Hua Zhou, Liang Diao, and Dongyan Wang. "ChineseFoodNet: A large-scale image dataset for chinese food recognition." arXiv preprint arXiv:1705.02743 (2017).

[15] Harashima, Jun, Yuichiro Someya, and Yohei Kikuta. "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, pp. 1229-1232. ACM, 2017.

[16] Marin, Javier, Aritro Biswas, Ferda Ofli, Nicholas Hynes, Amaia Salvador, Yusuf Aytar, Ingmar Weber, and Antonio Torralba. "Recipe1M: A Dataset for Learning Cross-Modal Embeddings for Cooking Recipes and Food Images." arXiv preprint arXiv:1810.06553 (2018).

[17] Bruno, Vieira, and Cui Juan Silva Resende. "A survey on automated food monitoring and dietary management systems." Journal of health & medical informatics 8, no. 3 (2017).

[18] Min, Weiqing, Shuqiang Jiang, Linhu Liu, Yong Rui, and Ramesh Jain. "A Survey on Food Computing." arXiv preprint arXiv:1808.07202 (2018).

[19] Gwo Giun Lee, Cheng-Liang Wang, He-Yuan Lin and Tihao Chiang, "An Automatic White Balancing Algorithm with Achromatic Color Detection," 2007 Symposium on Digital Life Technologies (SoDLT).

[20] Pizer, Stephen M., E. Philip Amburn, John D. Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart ter Haar Romeny, John B. Zimmerman, and Karel Zuiderveld. "Adaptive histogram equalization and its variations." Computer vision, graphics, and image processing 39, no. 3 (1987): 355-368.

Authorized licensed use limited to: Malta College of ArtsScience&Tech. Downloaded on May 27,2021 at 11:50:24 UTC from IEEE Xplore. Restrictions apply.