Cookpad Image Dataset: An Image Collection as Infrastructure for Food Research

Jun Harashima
Cookpad Inc.
Tokyo, Japan
jun-harashima@cookpad.com

Yuichiro Someya
Cookpad Inc.
Tokyo, Japan
yuichiro-someya@cookpad.com

Yohei Kikuta
Cookpad Inc.
Tokyo, Japan
yohei-kikuta@cookpad.com

ABSTRACT

In food-related services, image information is as important as text information for users. For example, in recipe search services, users find recipes based not only on text but also images. To promote studies on food images, many datasets have recently been published. However, they have the following three limitations: most of the datasets include only thousands of images, they only take account of images after cooking not during the cooking process, and the images are not linked to any recipes. In this study, we construct the Cookpad Image Dataset, a novel collection of food images taken from Cookpad, the largest recipe search service in the world. The dataset includes more than 1.64 million images after cooking, and it is the largest among existing datasets. Additionally, it includes more than 3.10 million images taken during the cooking process. To the best of our knowledge, there are no datasets that include such images. Furthermore, the dataset is designed to link to an existing recipe corpus and thus, a variety of recipe texts, such as the title, description, ingredients, and process, is available for each image. In this paper, we described our dataset's features in detail and compared it with existing datasets.

CCS CONCEPTS

• **Computing methodologies** → **Computer vision**; • **Information systems** → *Multimedia databases; Multimedia and multimodal retrieval*;

KEYWORDS

Image collection; food image; recipe

1 INTRODUCTION

Images play an important role in food-related services, for example, recipe search services such as Yummly,¹ restaurant search services such as Yelp,² and food delivery services such as UberEATS.³ To locate recipes, restaurants, and delivery services, users take account of not only text information but also image information.

¹www.yummly.co

²<https://www.yelp.com>

³<https://www.ubereats.com>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGIR '17, August 07-11, 2017, Shinjuku, Tokyo, Japan.

© 2017 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-5022-8/17/08.

<http://dx.doi.org/10.1145/3077136.3080686>

To promote studies on food images, many datasets have recently been published [1–8, 12, 15]; they can be obtained from their official home pages or by sending a request e-mail to their owners. These datasets have successfully contributed images to many studies, especially for food recognition and dietary assessment [10, 11, 13, 19].

However, existing datasets have the following three limitations: First, they do not include a wide range of food images. Most datasets include only thousands of images, and even Food-101 [2], which is one of the most popular datasets in this field, includes at most 101,000 images. This is not sufficient to construct complicated statistical models (e.g., recent models in deep learning) for food images.

Second, they do not take account of images during the cooking process (hereafter called *process images*); that is, they only take account of images after cooking (called *complete images*). To promote studies for a wide variety of novel research topics, such as cooking support systems based on motion recognition and recipe text generation from process images, not only complete images but also process images are indispensable.

Finally, images in existing datasets are not linked to any recipes. A recipe contains various texts, such as a title, description, ingredients, and process, and these texts are essential for many research topics, such as ingredient detection and calorie estimation [3, 10, 16, 18]. However, there have been very few efforts to focus on this aspect and thus, some researchers need to collect this text for themselves.

To solve these problems, we construct a novel image collection called the Cookpad Image Dataset, which consists of food images taken from Cookpad,⁴ the largest recipe search service in the world. The dataset is designed to have the following three features:

Large collection of complete images It includes more than 1.64 million complete images. This makes our dataset the largest among existing datasets.

Large collection of process images It also includes more than 3.10 million process images. To the best of our knowledge, there are no datasets that include such images.

Linkable recipe texts It can be connected with an existing recipe corpus, and various texts in the recipes is available for each image.

In the following sections, we introduce the existing datasets and then describe our dataset in detail.

⁴<https://cookpad.com>

2 RELATED WORK

Since approximately 2010, researchers have gradually paid more attention to image collections for food research. Here we describe some of the publicly available datasets in chronological order.

Pittsburgh fast-food image dataset (PFID) [4], proposed in 2009, is one of the earliest datasets for food research. As the name suggests, PFID focuses on fast-food images (e.g., burgers, pizza) and includes 4,545 images of 101 fast food items from 11 popular fast food chains.

In 2010, Rakuten Institute of Technology published Rakuten Data, which includes images and text of approximately 440,000 Japanese recipes [17]. With the exception of our dataset, this is now the only dataset whose images are linked to recipes. In 2016, the dataset expanded to approximately 880,000 recipes.

In 2012, Matsuda et al. proposed a new dataset called the University of Electro-Communications Food 100 (UEC FOOD 100) [15]. The dataset contains 9,060 images of 100 Japanese food categories. It has grown into UEC FOOD 256, which contains 31,397 images of 256 categories [12]. One notable characteristic of these datasets is that bounding boxes indicating the location of foods in each image are also available.

Chen et al. provide a food image collection on their research site. They study food identification and quantity estimation from images, and publish the dataset used in their experiments. In the dataset, there are 50 categories of Chinese foods and 100 images for each category [5].

In 2014, Farinella et al. proposed the University of Catania food dataset 889 (UNICT-FD889) [8]. The dataset includes 889 food categories; it is currently the dataset with the most categories. The dataset includes 3,583 images of those categories (four images per category on average).

Food-101 [2] is one of the most popular datasets in studies on food images [13, 14, 16, 19]. The dataset consists of 101 food categories (the top 101 most popular dishes in Foodspotting⁵) and each category consists of 1,000 images. Thus, Food-101 contains a wide coverage of 101,000 images.

In 2015, Beijbom et al. constructed Menu-Match, an image collection of restaurant-specific foods [1]. They collected 646 images, with 1,386 tagged items across 41 categories. Additionally, they provide calorie counts for all items given by a dietitian in the restaurants included in the dataset.

Ciocca et al. introduced the University of Milano-Bicocca 2015 (UNIMIB2015), which is composed of 2,000 tray images of multiple foods, and contains 15 categories [6]. They also introduced UNIMIB2016, which is composed of 1,027 images and 73 categories [7].

In 2016, VIREO Food-172 proposed a dataset specifically for Chinese dishes [3]. For each food category, images were crawled from Baidu and Google image search (a total of 110,241 images). The notable characteristic of this dataset is that 353 ingredients were annotated for the images by 10 homemakers.

These datasets are all publicly available and thus, have successfully contributed to food research. However, as described in the previous section, the limitations of the datasets are that they do not include a wide range of complete images, do not include any process images, and are not linked to any recipes.

⁵<http://www.foodspotting.com>

Figure 1: Example of a recipe from Cookpad.

3 COOKPAD IMAGE DATASET

In this paper, we introduce the Cookpad Image Dataset, a novel collection of food images collected from Cookpad. We describe the dataset in terms of the following three features: large collection of complete images, large collection of process images, and linkable recipe texts.

3.1 Large Collection of Complete Images

One of the most notable features of our dataset is that it includes a vast number of food images. We collected approximately 1.72 million Japanese recipes that had been uploaded to Cookpad by September 2014 and extracted all complete images from the recipes. Note that some recipes did not have a complete image; thus, we collected 1,606,537 complete images. Figure 1 provides an example of recipes in Cookpad, which consist of text (e.g., title, description, ingredients, and process) and images (complete and process images). The figure shows a recipe for ゴーヤチャンプル (gōyā chanpurū) and the upper red rectangle indicates the complete image of the recipe.

The dataset also includes many images for meals, such as combinations of multiple foods. We collected these meals from Cookpad and extracted 35,927 complete images. Figure 2 provides an

Figure 2: Example of a meal from Cookpad.

example of a meal from Cookpad. It is for ランチ (lunch) and it consists of two foods: チャーハン (fried rice) and サムギョブサル (samgyeopsal). In the figure, the upper red rectangle indicates the complete image for the meal. As described in the previous paragraph, the dataset also includes the complete images for each food, indicated by the lower red rectangles.

3.2 Large Collection of Process Images

The second notable feature of our dataset is that it includes a wide range of images of cooking processes. As seen in Figure 1, a recipe in Cookpad contains multiple processes, and each process consists of an image and text. Note that some processes do not have images. In the figure, the recipe contains four processes, and the lower red and gray rectangles represent the images and text, respectively. For example, the first process explains the preparation of the gōyā, such as removing the inner seeds. Similar to the complete images, we extracted all process images from the 1.72 million recipes and collected 3, 105, 594 images.

3.3 Linkable Recipe Texts

The third feature of our dataset is that a variety of text is available for each image. As described previously, the images in our dataset were extracted from recipes and meals that had been uploaded to Cookpad by September 2014. These are the same recipes and meals targeted in Harashima’s text corpus [9], which is the largest corpus in food research. Thus, it is possible that the images in our dataset can be connected with the text in the corpus.

Consider again Figure 1. Harashima’s corpus includes the title, description, ingredients, and process, indicated by gray rectangles. Note that although the figure only shows the main information for the recipe, it also includes the unique ID of the recipe, unique ID of the author, categories of the recipe (if any), and date when the recipe was uploaded. Using our dataset with the corpus, researchers can utilize not only images but also text for their studies.

Similar to recipes, a variety of text can be used with images for each meal. In Figure 2, the upper gray rectangles represent the title, noteworthy points, and cooking time, and this information is also available from Harashima’s corpus. Additionally, the corpus includes the unique ID of the meal, unique ID of the author, categories of the meal (if any), and date when the meal was uploaded. Using the two datasets, we can use both images and text for various studies.

4 DISCUSSION

In Table 1, we summarize the statistics and notable features of existing datasets and our dataset. From the third column in the table (# of complete images), we can see that our dataset includes the most complete images among all datasets. Most existing datasets include only thousands of images, and even Rakuten Data, Food-101, and VIREO Food-172 include at most hundreds of thousands of images. By contrast, our dataset consists of more than 1.64 million images, which includes approximately 1.61 million images of single foods and approximately 36, 000 images of multiple foods. It is preferable that more data is used in many studies, especially for machine learning.

As seen in the fourth column of Table 1, the Cookpad image dataset is the only dataset that stores not only complete images but also process images. We believe that these images promote a wide variety of novel research topics. For example, researchers can use them to study cooking support systems that recognize and support cooking processes, such as reading process text in accordance with the situation. Another example is that the images can be used in studies for text generation, such as generating a recipe from process images.

The last column shows that there are no datasets whose images can be linked to recipe text, except for Rakuten Data, which is similar to our dataset regarding this feature. However, although Rakuten Data includes only complete images of single foods, our dataset also includes complete images of multiple foods and process images; thus, it can also connect those images with recipe text.

By contrast, images in VIREO Food-172 are linked to 353 ingredients. However, images in our dataset can be connected with not only ingredients in Harashima’s corpus but also other text in the corpus, such as the title, description, and process.

Table 1: Statistics and features of existing datasets and our dataset.

	year	# of complete images	# of process images	notable features
PFID [4]	2009	4, 545	N/A	fast-food
Rakuten Data [17]	2010	approx. 800, 000	N/A	linkable recipe texts
UEC FOOD 100 [15]	2012	9, 060	N/A	100 categories, bounding boxes
Chen's dataset [5]	2012	5, 000	N/A	50 categories
UEC FOOD 256 [12]	2014	31, 397	N/A	256 categories, bounding boxes
UNICT-FD889 [8]	2014	3, 583	N/A	889 categories
Food-101 [2]	2014	101, 000	N/A	101 categories
Menu-Match [1]	2015	646	N/A	restaurant food, 41 categories, calorie counts
UNIMB2015 [6]	2015	2, 000	N/A	15 categories
UNIMB2016 [7]	2016	1, 027	N/A	73 categories
VIREO Food-172 [3]	2016	110, 241	N/A	172 categories, 353 ingredients
Cookpad Image Dataset	2017	1, 642, 450	3, 105, 594	large collection of complete and process images, linkable recipe texts

Similar to most existing datasets, our dataset can also use category information through Harashima's corpus, which includes approximately 1, 100 categories (e.g., meat dishes, seafood dishes, and vegetable dishes). This is more than UNICT-FOOD889, which has the most categories among existing datasets.

Although our dataset has the aforementioned advantages over existing datasets, it has also two disadvantages. First, unlike UEC FOOD 100 and 256, our dataset does not have bounding box information that indicates the location of foods in each image. Second, unlike Menu-Match, our dataset does not have a calorie count for each image. Both types of information are useful for studies, especially for food recognition and calorie estimation. We will consider annotation in a future study.

5 CONCLUSIONS

In this paper, we introduced a novel image collection for food research. The dataset not only stores more than 1.64 million complete images, which includes approximately 1.61 million images of single foods and approximately 36,000 images of multiple foods, but also more than 3.10 million process images. This is now the largest food image dataset in the world. Using an existing corpus, researchers can also exploit recipe text, such as the title, description, ingredients, and process, for each image. We believe that these features of our dataset promote a variety of studies on food images. The dataset can be obtained by sending an e-mail request to the authors. In future work, we plan to annotate the images in our dataset with further information, such as bounding boxes and calorie counts.

REFERENCES

- [1] Oscar Beijbom, Neel Joshi, Dan Morris, Scott Saponas, and Siddharth Khullar. 2015. Menu-Match: Restaurant-Specific Food Logging from Images. In *Proceedings of the 2015 IEEE Winter Conference on Applications of Computer Vision (WACV 2015)*. 844–851.
- [2] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. 2014. Food-101 – Mining Discriminative Components with Random Forests. In *Proceedings of the 13th European Conference on Computer Vision (ECCV 2014)*. 446–461.
- [3] Jingjing Chen and Chong-Wah Ngo. 2016. Deep-based Ingredient Recognition for Cooking Recipe Retrieval. In *Proceedings of the 2016 ACM on Multimedia Conference (ACMMM 2016)*. 32–41.
- [4] Mei Chen, Kapil Dhingra, Wen Wu, Lei Yang, Rahul Sukthankar, and Jie Yang. 2009. PFID: Pittsburgh Fast-Food Image Dataset. In *Proceedings of the 16th IEEE International Conference on Image Processing (ICIP 2009)*. 289–292.
- [5] 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 *Proceedings of the 5th ACM SIGGRAPH Conference and Exhibition on Computer Graphics and Interactive Techniques in Asia (SIGGRAPH Asia 2012)*.
- [6] Gianluigi Ciocca, Paolo Napoletano, and Raimondo Schettini. 2015. Food Recognition and Leftover Estimation for Daily Diet Monitoring. In *New Trends in Image Analysis and Processing – ICIAP 2015 Workshops*. 334–341.
- [7] Gianluigi Ciocca, Paolo Napoletano, and Raimondo Schettini. 2017. Food Recognition: a New Dataset, Experiments and Results. *IEEE Journal of Biomedical and Health Informatics* (2017).
- [8] Giovanni Maria Farinella, Dario Allegra, and Filippo Stanco. 2014. A Benchmark Dataset to Study the Representation of Food Images. In *Proceedings of the 13th ECCV Workshop on Assistive Computer Vision and Robotics (ACVR 2014)*. 584–599.
- [9] Jun Harashima, Michiaki Ariga, Kenta Murata, and Masayuki Ioki. 2016. A Large-Scale Recipe and Meal Data Collection as Infrastructure for Food Research. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*. 2455–2459.
- [10] Hongsheng He, Fanyu Kong, and Jindong Tan. 2016. DietCam: Multi-View Food Recognition Using a Multi-Kernel SVM. *IEEE Journal of Biomedical and Health Informatics* 20, 3 (2016), 848–855.
- [11] 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*. 350–357.
- [12] Yoshiyuki Kawano and Keiji Yanai. 2014. Automatic Expansion of a Food Image Dataset Leveraging Existing Categories with Domain Adaptation. In *Proceedings of the 13th ECCV Workshop on Transferring and Adapting Source Knowledge in Computer Vision (TASK-CV 2014)*. 3–17.
- [13] Chang Liu, Yu Cao, Yan Luo, Guanling Chen, Vinod Vokkarane, and Yunsheng Ma. 2016. DeepFood: Deep Learning-Based Food Image Recognition for Computer-Aided Dietary Assessment. In *Proceedings of the 14th International Conference on Inclusive Smart Cities and Digital Health (ICOST 2016)*. 37–48.
- [14] Jonathan Malmaud, Jonathan Huang, Vivek Rathod, Nick Johnston, Andrew Rabinovich, and Kevin Murphy. 2015. What's Cookin' ? Interpreting Cooking Videos using Text, Speech and Vision. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2015)*. 143–152.
- [15] 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)*. 25–30.
- [16] Austin Myers, Nick Johnston, Vivek Rathod, Anoop Korattikara, and Alex Gorban. 2015. Im2Calories: towards an automated mobile vision food diary. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV 2015)*. 1233–1241.
- [17] Rakuten Institute of Technology. 2010. Rakuten Data Release. (2010). <http://rit.rakuten.co.jp/opendata.html>.
- [18] Kyoko Sudo, Jun Shimamura, Kazuhiko Murasaki, and Yukinobu Taniguchi. 2014. Estimating nutritional value from food images based on semantic segmentation. In *Proceedings of the Workshop on Smart Technology for Cooking Eating Activities (CEA 2014)*. 571–576.
- [19] Keiji Yanai and Yoshiyuki Kawano. 2015. Food Image Recognition using Deep Convolutional Network with Pre-training and Fine-tuning. In *Proceedings of the 7th Workshop on Multimedia for Cooking and Eating Activities (CEA 2015)*. 1–6.