

Food Ingredients Identification from Dish Images by Deep Learning

Ziyi Zhu, Ying Dai

Faculty of Software and Information Science, Iwate Prefectural University, Iwate, Japan Email: g231r015@s.iwate-pu.ac.jp, dai@iwate-pu.ac.jp

How to cite this paper: Zhu, Z.Y. and Dai, Y. (2021) Food Ingredients Identification from Dish Images by Deep Learning. *Journal of Computer and Communications*, **9**, 85-101. https://doi.org/10.4236/jcc.2021.94006

Received: March 1, 2021 **Accepted:** April 22, 2021 **Published:** April 25, 2021

Copyright © 2021 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). http://creativecommons.org/licenses/by/4.0/

<u>.</u>

Open Access

Abstract

Diet plays an important role in people's daily life with its strong correlation to health and chronic diseases. Meanwhile, deep based food computing emerges to provide lots of works which including food recognition, food retrieval, and food recommendation, and so on. This work focuses on the food recognition, specially, the ingredients identification from food images. The paper proposes two types of ways for ingredient identification. Type1 method involves the combination of salient ingredients classifier with salient ingredient identifiers. Type 2 method introduces the segment-based classifier. Furthermore, this work chooses 35 kinds of ingredients in the daily life as identification categories, and constructs three kinds of novel datasets for establishing the ingredient identification models. All of the classifiers and identifiers are trained on Resnet50 by transfer learning. Many experiments are conducted to analyze the effectiveness of proposed methods. As the results, Salient ingredients classifier predict one ingredient and achieves 91.97% on test set of salient ingredients dataset and 82.48% on test dish image dataset. Salient ingredients identifiers predict remained ingredients and achieve mean accuracy of 85.96% on test dish image dataset. Furthermore, Segment-based classifier achieves 94.81% on test set of segment-based ingredients dataset.

Keywords

Ingredient Identification, Dataset, Salient Ingredient, Segmentation, Classification

1. Introduction

In recent years, people are more and more concern about diet health since a healthy diet helps prevent all types of malnutrition disease, including diabetes, heart disease, stroke and cancer. According to the World Health Organization's food health standard, it is a way to help judge food health by recording the kinds of food ingredients consumed every day. At the same time, with the development of AI [1], a food recognition field emerged, which contains a series of tasks such as automatic identification of food names and food recipes from food images through AI. In this way, there are many works that are proposed which aim to replace traditional handwritten records.

With the rapid development of the social network, much food-related information appears every day, including recipe sharing, cooking videos, and diet tracking. Under this background, food computing [2] is introduced for many kinds of food-related research and is emerging as a new field to address the issues from many food-relevant fields, such as nutrition, agriculture, and medicine. As significant tasks in food computing, Food Recognition [3] [4] [5] which recognize food categories from test food images, and Food Retrieval [6] [7] which query the most similar food image from database have received more attention in multimedia. In this research, the authors focus on food Recognition, especially, the food ingredient identification.

Ingredients are important components of food, and the performance of food recognition can be improved by introducing the ingredient information as complement to resolve the high intra-class variations by same food category. Under this background, in this research, authors propose a novel system of food ingredients identification which involves in two types of ingredient identification methods: 1) combining salient ingredient classifier with salient ingredients identifiers; 2) constructing segment-based ingredients classifier. All of these classifiers and identifiers are trained on Resnet50 by transfer learning.

For type1 method, the dish image is inputted into salient ingredients classifier to predict one salient ingredient. Salient ingredients identifiers are used to predict other ingredients of the dish image. Identifiers are chosen according to the ingredients which are searched from the ingredients co-occurrence matrix. For type2 method, each ingredient is extracted from the dish image, and segment-based classifier identifies these ingredients in the dish image.

Furthermore, the authors propose a novel hierarchical ingredient dataset structure based on MAFF policies of food quality labelling standard [8] and MAFF of crop classification [9]. According to this structure, 35 ingredient categories in the daily life are chosen to construct two types of datasets for training the proposed models. Salient ingredients dataset includes total 6662 images labelled by 35 categories of ingredients; segment-based ingredients dataset includes total 7803 segmented ingredient images labelled by the same 35 ingredients categories. Furthermore, the test dish image dataset is constructed to evaluate type1 method, and it contains dish images which not overlap with Salient ingredients dataset. Each dish image in this dataset contains single or multiple ingredients with or without salient ingredient inside.

The authors conducted extensive experiments to verify the effectiveness of the proposed methods. For the salient ingredients classifier, the prediction accuracy of salient ingredient reaches 91.97% on the salient ingredients test dataset. For

the segment-based ingredients classifier, the accuracy reaches 94.81% on the segment-based ingredients test dataset. For the salient ingredient identifiers, the mean average accuracy reaches 85.96% on the test dish image dataset.

The authors further investigate the Grad-CAM [10] visualization results of the classifiers, and find that three reasons cause the misclassification: 1) interference by other ingredients; 2) high inter-class similarity; 3) high intra-class variations. For segment-based ingredients classifier, the interference by other ingredients is solved. However, the other two types of misclassification are unsolved.

The reminder of the paper is organized as follows: Section 2 gives some related works regarding food and ingredient recognition, and indicates the issues in the current research; Section 3 introduces the construction of food ingredient datasets, which are used to train the ingredient identification models; Section 4 explains the methods proposed for ingredient identification; Section 5 provides the experimental results and analyze the effectives of the proposed methods. Section 6 concludes this paper point out key points and future direction.

2. Related Work

2.1. Food Recognition

One commonly task of food computing is the food recognition from images. Food recognition can be widely used in many fields such as dietary tracking, food recommendation, cooking learning from the food images. Furthermore, food recognition can even apply to commercial scenarios such as automatic purchasing system, automatic checkout at supermarkets or restaurants.

Unfortunately, it is tough to capture the discriminative features for food recognition, because food items are deformable and exhibit significant intra-class variations and inter-class similarity in appearance. Multi-Task Learning for Food Identification and Analysis with Deep Convolutional Neural Networks [11] proposed a multi-task system for simultaneously training a dish identifier, a cooking method recognizer and a multi-label ingredient detector. Ingredient-Guided Cascaded Multi-Attention Network for Food Recognition [4] achieves food recognition by developing an Ingredient-Guided Cascaded Multi-Attention Network (IG-CMAN), which sequentially localizes multiple informative image regions with multi-scale from category-level to ingredient-level guidance. Furthermore, Multi-Scale Multi-View Deep Feature Aggregation for Food Recognition [3] proposes a multi-scale multi-view feature aggregation scheme for food recognition from multi-view and multi-scale of food image.

Although all these works recognize ingredients in a certain stage and achieve plausible accuracy on food recognition, but they do not focus to resolve the problems of the ingredient classification: 1) high intra-class variants; 2) high inter-class similarity; 3) interference by other ingredient. In addition, their ingredient categories are not strictly defined in common sense. For example, they serve potato slice and potato stick as different categories but in fact they both belong to potato ingredient just with different appearances due to the different cutting method.

2.2. Food Ingredient Identification

Although tens of thousands of dishes exist, they are composed of a much smaller number of ingredients. Accordingly, ingredients can be served as a subclass of food recognition. However, ingredient recognition receives much less attention than food recognition. Some works based on deep models have recently started to focus on ingredient recognition to improve food recognition performance. Food Ingredients Recognition Though Multi-label Learning [12] propose a model which involve multi-label learning to identify ingredients. Zero-Shot Ingredient Recognition by Multi-Relational Graph Convolutional Network [13] proposed a framework which is composed of two modules: a multi-label deep convolutional neural network (DCNN) for known ingredient classifier learning and a multi-relational graph convolutional network (mRGCN) for unseen ingredient prediction. Even though lots of works are proposed to resolve food recognition with additional ingredients attributes, few works focus on addressing the problems for ingredients identification, which are identifying ingredients with high intra-class variation and high inter-class similarity. Therefore, this work proposes a novel mechanism which focuses on addressing the abovementioned problems by introducing additional supervised information from salient and segmented ingredients for food ingredient identification.

3. Dataset Construction

3.1. Ingredient Categories Structure

The biological-based food ingredient dataset structure is important since the high inter-class similarity often comes from the biological similarity. However, there is no food dataset construction based on the biological attributes of the ingredients and unique for the ingredient recognition task. Under the above requirements, this work proposes a tree structure for constructing the ingredients dataset based on MAFF policies of food quality labelling standard [8] and MAFF of crop classification [9]. The different biological levels of ingredients are hierarchically defined, which are shown as **Figure 1**. By the way, this tree structure can be extended by other ingredients attributes, such as cooking and cutting methods. In addition, this structure can also be served as a bridge that can unified existing datasets to address the problem of limited food coverage. The unified datasets aim to handle tasks in the food computing area by selecting experimental resources from different tree depths. For example, the diet tracking task needs data samples of level 2; the food recognition needs data samples of level 3; cooking learning needs data samples of level 4.

However, in order to construct the real-world dataset to train the ingredient recognition models and validate the proposed methods, only 35 common ingredient categories in daily life are selected in this work. Figure 2 shows these 35 ingredient categories.



Figure 1. Ingredients tree.



Figure 2. 35 ingredients.

3.2. Datasets

This work involves in three kinks of datasets. To construct these datasets, first select food images from xiachufang.com with the queries according to 35 ingredient categories in **Figure 2**. All food images include ingredients which can be distinguished by human vision. And then annotate each of the dataset by different labels for various training purposes. For setting of train-test samples, all datasets are split into 70%, 10%, 20% images for training, validation and testing, respectively. Next, more construction details of each dataset will be explained.

• Salient Ingredients dataset (SAIng)

This dataset aims to train the salient ingredients classifier and salient ingredient identifiers. At first, food images dataset which contain salient ingredient images according to each ingredient, respectively. However, at present, no dataset can help us to train and verify the proposed models of ingredients identification. For construction, each food image contains one salient ingredient inside dish regions are extracted from all selected food images by trained dish detection model and then label the each dish image by salient ingredient. This dataset also aims to train model to learn more discriminative features for each ingredient then other dataset with multiple ingredients inside. Especially, this work collects dish images with its salient ingredient and has different appearances for each ingredient category. For the statistic, the dataset contains 35 types of ingredients categories with 6662 images. The distribution of this dataset shows in **Figure 3**, and **Table 1** shows some data samples and corresponding ingredients name from the salient ingredients dataset.



Figure 3. Distribution of sample number of all ingredient categories in salient ingredient set.



Table 1. Examples of images from salient based ingredients dataset.

• Segment-based ingredients dataset (SEIng)

This dataset keeps the same amounts of categories as the salient dataset and contains 7803 segmented ingredient images. For the preparation of data samples. Before segmentation of ingredients from dish images, this work first extract dishes from food images. The trained Dish Detection model based on Faster RCNN can be used to generate dish images. Then for ingredients segmentation, first calculate Super-pixels of the image based on L * a * b color space. Then use k-means clustering-based image segmentation to output each cluster of segment ingredients where k = ingredients number. Then, manually select useful segmentation results which can represent each ingredient from all segmentation results. Furthermore, in each ingredient category, this dataset makes sure data samples include as many appearance variations as possible, including different cutting styles and cooking methods. For example, potato slice, potato stick, and boiled potato. This dataset's distribution shows in **Figure 4** and **Table 2** shows some data samples from this dataset.

• Test dish image dataset

To verify the effectiveness of salient ingredient classifier and salient ingredient identifiers in identifying the salient ingredient from multiple ingredient images, this work constructs a test dish image dataset. This dataset contains dish images with multiple ingredients inside and not overlap the data samples with Salient ingredients dataset and Segment-based ingredients dataset. Meanwhile, most dish images contain 1 - 4 ingredients inside per dish image, and few samples contain more than 4 ingredients. **Table 3** shows the examples of dish images with different number of ingredients.







Figure 4. Distribution of sample number of all ingredient categories in segment-based ingredient set.

 Table 3. Examples of dish images contain different number of ingredients.



4. Ingredient Identification

The goal of this work is to identify visible ingredients from food images. The ingredient identification model is constructed for identifying the ingredients inside the dish. Accordingly, this paper proposes two methods for this goal: 1) Combination of Salient ingredient classifier (A1) with Salient ingredients identifiers (Bs) by ingredients co-occurrence matrix; 2) Segment-based ingredient classifier (A2). Note that, by using this method, multi-label ingredients classification can be exchanged to multi-class ingredients classification. The detailed introduction is explained as the following.

4.1. Problem Analysis

Ingredient recognition is not an easy task. Difficulties of ingredients recognition come from these aspects: 1) Different cutting and cooking methods can make the same ingredients look quite differently, however, different ingredients look quite similar. 2) Other Ingredients interference: dish images contain multiple ingredients and are difficult to focus on specific ingredients when training since the frequent appearance of some kind of ingredients composition may influence learning features. In order to improve the recognition performance for ingredients, some works puts forward the use of auxiliary attribute information. Likes, cooking attributes, cutting attributes for ingredient identification. Our method is proposed to settle the problems of high intra-class variations of ingredient and high inter-class similarity of ingredients by training salient ingredient classifier with salient ingredient supervised and also segmented ingredient classifier with segmented ingredient supervised to find more details of ingredients features. For more details, this work proposes the Salient ingredient method to identify one ingredient from dish image and to verify whether detail attributes of ingredients can be learned effectively, and also proposes an ingredient segmentation method which provides a new way to identify food ingredients from segmented ingredient images. These two methods have high research value but with high challenges for food ingredients Recognition. The research diagram in this paper is shown in **Figure 5**.

4.2. Salient Ingredient Classifier (A1)

This classifier is trained to identify the main ingredient in the dish image. This classifier is trained by the dataset SAIng. This means there is only one salient in the image, and other ingredients should be seen as background.

4.3. Ingredient Identifiers (Bs)

Because most dish is the composition of multiple ingredients, so the ingredient identification can be seen as multi-label classification. The simplest way to resolve this task is to transform multi-label classification into single-label classification by training multiple binary identifiers for each category. In this work, identifiers are trained for each ingredient by the dataset SAIng, respectively, so as to generate 35 identifiers called salient ingredient identifiers (Bs).

4.4. Combination of A1 and Bs

This work achieves the multiple ingredients identification in the dish image by





leveraging ingredients co-occurrence relationship. 1) Construct the ingredients co-occurrence matrix based on a large-scale dataset which contains multi-label information of each data sample; 2) Predict salient ingredient from food image by salient ingredients classifier and use salient ingredient as a reference to search co-occurrence ingredients from co-occurrence matrix and rank these ingredients by co-occurrence values to generate sorted co-ingredients list. 3) Use ingredient counting model [14] to predict ingredient number of dish image; 4) Identify the ingredients successively in the order of sorted co-ingredients list until reach the ingredient number.

4.5. Segment-Based Ingredients Classifier (A2)

According to the intuitive insights of human visual, people always need to determine each ingredient's region from the food image, before they identify ingredient category. Therefore, the crucial point of food ingredient identification is to detect the region of each food ingredient.

Similar to scene segmentation, the range of food in the image is often disconnected. But the difference is that the shape of a different region of each ingredient is irregular. Therefore, the detection of the ingredient region is more complicated. This work aims to make the classifier focus on a single ingredient region to learn more distinctive features of ingredients during the training process. Furthermore, training on segmented ingredient regions can remove the interference by other ingredients when prediction. Based on this idea, this work proposes a method of inputting extracted dish images into ingredients counting model to predict the number of ingredients inside the dish images. Then, K-means is utilized to segment each ingredient according to number of ingredients in the dish image. Finally, each of segmented ingredient images is inputted into segment-based classifier to predict ingredient categories. Segment-based ingredients classifier is trained by the dataset SEIng.

In this way, the original multi-label ingredient classification can be transformed to single-label ingredient classification.

5. Experiments Results and Analysis

The ingredient identification involves in two methods, A1 + B and A2. A1, B, A2 are constructed by transfer learning Resnet50 pre-trained model on SAIng and SEIng dataset respectively. For training samples, the 4/5 of samples in SAIng or SEIng are randomly chosen. For validation samples, the remained 1/5 samples in SAIng or SEIng are chosen. In order to verify the capability of A1 and Bs on identifying multiple ingredients image, A1 and Bs are further tested on test dish image dataset. Experiments are implemented on Matlab environment.

5.1. Evaluation Metrics

1) Accuracy

This work uses accuracy to evaluate all my models and accuracy is defined as follows:

$$Accuracy = \frac{Correct \ Predictions}{Total \ Number \ of \ Examples}$$
(1)

2) Precision and Recall

Since accuracy is inappropriate for imbalance classification. Precision and recall [15] are used for imbalance classification. Precision summarizes the fraction of samples are predicted as positive class that belong to the positive class. Its formula is shown as follow:

$$Precision = \frac{True Positive}{True Positive + False Positive}$$
(2)

Recall summarizes the fraction of samples are well predicted that belong to the positive class. Its formula is shown as follow:

$$Recall = \frac{True Positive}{True Positive + False Negative}$$
(3)

5.2. Experimental Results for A1

A1 model is evaluated on validation set of SAIng, and analyze the precision and recall values corresponding to each ingredient categories. The precision and recall value are shown in **Figure 6**. From this result, the results show that 77% of ingredients achieve precision and recall over 80%.



Figure 6. Precision and Recall value of A1.

In order to verify the effectiveness of A1 on multi-ingredients food images, authors further evaluate A1 model on test dish image dataset. And accuracy is defined by:

$$\operatorname{accuracy} = \frac{\operatorname{sum}(Y_i \text{ is in } Ing_i)}{\operatorname{total images}}$$
(4)

 Ing_i is the multi-ingredients label vector of each dish image. And Y_i is the prediction by A1 model. The accuracy achieves 82.48%. Compare with accuracy results test on SAIng, A1 model is able to identify salient ingredient from dish image with multiple ingredients inside even without salient ingredient inside.

Visualization of prediction results and are shown in Table 4, from this table, activated regions of many samples (highlighted in warm colors) are semantically meaningful. And from the results, the model seems to have the ability to exclude irrelevant ingredient regions from recognition. The first row shows that the model misclassifies the dish images due to the interference by other ingredients. The second row shows that fried egg is misclassified since SAIng contains few training samples of fried egg than scramble egg. Hence, Imbalance distribution of training samples causes one type of misclassification and can be partially resolved by adding more data samples with different appearance of each ingredient. The third row shows the examples of ingredients with the high inter-class similarity. Cauliflower is predicted by egg, and garlic stem is predicted by asparagus since they have similar appearance with misclassified ingredient respectively. For conclusion, three types of misclassification by A1 can be summarized.

- Interference by other ingredients.
- Imbalance Distribution of training samples.
- High inter-class similarity.

Error type Samples (L: original image R: Grad CAM visualization) auliflower

Table 4. Examples of grad cam visualization from the misclassification of A1.

Fruite elon (0.76) Interference by other ingredients Imbalance distribution of training samples High inter-class similarity

DOI: 10.4236/jcc.2021.94006

5.3. Experimental Results for Bs

Bs are tested on test set of SAIng. The precision and recall values are shown in **Figure 7**. The average precision and recall of B research 89.59%, 94.07% respectively. In addition, 77% of ingredients reach precision and recall value over 80%. Moreover, some interesting findings we observed are 1) some ingredients with both high precision and recall values have distinctive visual appearance, like corn, broccoli; 2) Ingredients with high precision but low recall value have high intra-class variations like yam. Pumpkin; 3) Ingredients with high recall but low precision has high inter-class similarity like asparagus, onion. 4) Ingredients with both low precision and recall are caused by high inter-class similarity and intra-class variations at the same time.

In order to verify the effectiveness of Bs on multi-ingredients food images, more experiments are conducted on test dish image dataset. The accuracy is defined by Equation (4) whereas Y_i is the prediction by B model. The accuracy results of each ingredient type are shown in **Figure 8**. From the results, Bs perform well even though the food image without salient ingredient and the mean accuracy achieves 85.96%, and 77% of ingredients reach precision and recall value exceed 80%.

5.4. Experimental Evaluation for A2

A2 is tested on the validation set of SEIng. And the results of precision and recall value are shown in **Figure 9**. The average precision of A2 is 97.52%, and the average recall achieves 95.23%, and 88% of ingredients reach both precision and recall value over 80%.







Figure 8. Accuracy of B on test dataset.



Figure 9. Precision and recall value of A2.

DOI: 10.4236/jcc.2021.94006

| Test image | Wrong prediction | Ground Truth |
|------------|------------------|--------------|
| | Fruits elon | celery |
| | bean prout | enoki |
| | egg | cauliflower |

Table 5. Examples of grad cam visualization from the misclassification of A2.

To understand what the model is able to learn from the data samples, the visualization of the network by Grad-CAM is used to interpret whether the network is able to learn discriminative feature for each category, so as to to verify the effectiveness of A2 more justifiably.

By analyzing the misclassification case of A2, it is found that the inference of the background on the classification is eliminated, although the influences of the high inter-class similarity on the classification are still remained. On the other hand, the imbalance Distribution of training samples still affects the performance of A2.

To provide further insights, we demonstrate some wrong misclassification samples and show in **Table 5**. The results show that even wrong prediction samples activated from the region of the ingredients and are semantically meaningful. Furthermore, the more fine-grained features of ingredients are necessary for the identification of the ingredients with high inter-class similarity and high intra-class variations.

6. Conclusions

In this work, two kinds of methods are proposed for ingredients identification: 1) Combination of Salient ingredients classifier (A1) and Salient ingredients identifiers (Bs) by ingredient co-occurrence matrix; 2) Segment-based ingredients classifier (A2). Experimental results on correspond test dataset show that A1 model can identify one ingredient from multiple ingredients dish images. However, since A1 model can only predict one ingredient, the remained ingredients need Bs to identify. Moreover, for combination of A1 with Bs, the ingredients co-occurrence matrix is needed, which hasn't been constructed. However, it can be constructed by using recipes information in the future work. Utilizing all of Bs can identify all ingredients, but is not efficient.

A2 can identify all ingredients if the ingredients are extracted from the multi-ingredients food images. Moreover, A2 outperforms Bs since 88% of ingredient categories reach precision and recall value over 80%, where Bs only have 77% of ingredient categories. Moreover, A2 classifies the segmented ingredient image, which leads to transformation of multi-label classification to single-label classification. So the uncertainty of prediction number can be solved. Moreover, ingredient segmentation improves the performance of classification because of removing the interference of other ingredients. In order to segment each ingredient from food images sufficiently, we need further research on the segmentation method.

On the other hand, from the experimental results, high-intra class variations can be partially solved by adding many training samples with different shapes. However, the issue of high inter-class similarity and high intra-class variation of ingredient needs to be explored further. Moreover, some abovementioned interesting findings need to be researched detailedly in the next work.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/CVPR.2016.90
- [2] Min, W., Jiang, S., Liu, L., Rui, Y. and Jain, R.C. (2019) A Survey on Food Computing. ACM Computing Surveys (CSUR), 52, 1-36. <u>https://doi.org/10.1145/3329168</u>
- Jiang, S., Min, W., Liu, L. and Luo, Z. (2020) Multi-Scale Multi-View Deep Feature Aggregation for Food Recognition. *IEEE Transactions on Image Processing*, 29, 265-276. <u>https://doi.org/10.1109/TIP.2019.2929447</u>
- [4] Min, W., Liu, L., Luo, Z. and Jiang, S. (2019) Ingredient-Guided Cascaded Multi-Attention Network for Food Recognition. *Proceedings of the 27 th ACM International Conference on Multimedia*, Nice, 21-25 October 2019, 1331-1339. <u>https://doi.org/10.1145/3343031.3350948</u>
- [5] Pehlic, A., Almisreb, A., Kunovac, M., Skopljak, E. and Begovic, M. (2019) Deep Transfer Learning for Food Recognition. *Southeast Europe Journal of Soft Computing*, 8. <u>https://doi.org/10.21533/scjournal.v8i2.182</u>
- [6] Cao, D., Yu, Z., Zhang, H., Fang, J., Nie, L. and Tian, Q. (2019) Video-Based

Cross-Modal Recipe Retrieval. *Proceedings of the 27th ACM International Conference on Multimedia*, Nice, 21-25 October 2019, 1685-1693. https://doi.org/10.1145/3343031.3351067

- [7] Fontanellaz, M., Christodoulidis, S. and Mougiakakou, S. (2019) Self-Attention and Ingredient-Attention Based Model for Recipe Retrieval from Image Queries. *Proceedings of the 5th International Workshop on Multimedia Assisted Dietary Management*, Nice, 21-25 October 2019, 25-31. https://doi.org/10.1145/3347448.3357163
- [8] MAFF Policies of Food Quality Labelling Standard. <u>https://www.caa.go.jp/policies/policy/food_labeling/quality/quality_labelling_stand</u> <u>ard/pdf/kijun_01.pdf</u>
- [9] MAFF of Crop Classification. https://www.maff.go.jp/j/nouyaku/n_sasshin/group/sakumotu_bunrui.html
- [10] Selvaraju, R.R., Das, A., Vedantam, R., Cogswell, M., Parikh, D. and Batra, D. (2019) Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *International Journal of Computer Vision*, **128**, 336-359. https://doi.org/10.1007/s11263-019-01228-7
- [11] Zhang, X., Lu, Y. and Zhang, S. (2016) Multi-Task Learning for Food Identification and Analysis with Deep Convolutional Neural Networks. *Journal of Computer Science and Technology*, **31**, 489-500. <u>https://doi.org/10.1007/s11390-016-1642-6</u>
- Bolaños, M., Ferrà, A. and Radeva, P. (2017) Food Ingredients Recognition through Multi-Label Learning. *International Conference on Image Analysis and Processing*, Catania, 11-15 September 2017, 394-402. https://doi.org/10.1007/978-3-319-70742-6_37
- [13] Chen, J., Pan, L., Wei, Z., Wang, X., Ngo, C. and Chua, T. (2020) Zero-Shot Ingredient Recognition by Multi-Relational Graph Convolutional Network. *Proceedings* of the AAAI Conference on Artificial Intelligence, 34, 10542-10550. https://doi.org/10.1609/aaai.v34i07.6626
- [14] Tainaka, S. and Ying, D. (2020) Construction of a Model for Identifying Ingredients of Single Dishes Using CNN. *Proceedings of the 83rd National Convention of IPSJ*, Kanazawa, March 2020.
- [15] Classification: Precision and Recall. <u>https://developers.google.com/machine-learning/crash-course/classification/precisi</u> <u>on-and-recall</u>