Training Data Augmentation for Semantic Segmentation of Food Images Using Deep Learning

Takayuki Yamabe
Department of mechanical engineering, Kanazawa University, Address
Kanazawa, Ishikawa, Japan

Tatsuya Ishichi
Tokuo Tsuji
Tatsuhiro Hiramitsu
Hiroaki Seki

E-mail: yamabetakayuki@stu.kanazawa-u.ac.jp
http://as.ms.t.kanazawa-u.ac.jp

Abstract

We propose a training data generation method for semantic segmentation of food ingredients. Training data for semantic segmentation requires a large amount of effort for a human to carefully paint the boundaries of objects. Therefore, we propose a method to automatically augment appropriate training data by adding image composition processes for images of only one type of food. In our experiments, we confirmed the effectiveness of each data augmentation.

Keywords: data augmentation, food recognition, semantic segmentation, deep learning

1. Introduction

Food recognition from image is in demand for health care and food product quality control. The recognition of general food categories has reached practical accuracy and is being used in several applications[1]. On the other hand, the recognition of quantitative balance of food ingredients is an unresolved issue. One of the methods for recognizing food balance is semantic segmentation based on deep learning, which we focus on in this study. Semantic segmentation is a process that determines the region of an object in pixels for an image.

In order to achieve highly accurate image recognition, it is necessary to have a large amount of training data which is generated by human hand. In response to this, there has been research into making data creation more efficient [2], and research into data expansion, which ensures data diversity by transforming a small number of data [3]. In this study, we propose a training data generation method that is effective for semantic segmentation of food ingredients.

We define an image of mixed food as an image of several kinds of food, and an image of one kind food as an image of only one kind of food. For food balance recognition, we need to identify the images of mixed food, and the images needed for training are also images of mixed food. The labeling of images of mixed food requires manual and careful painting of the boundaries of the food ingredients, which is a large labor-intensive process. On the other hand, the images of one kind food can be labeled automatically based on the difference in color between the background and the food ingredients. Therefore, the labeling process can be easier if images of one kind food are combined to produce composite images like images of mixed food. This method of generating training data using composite images has been used in binarization tasks such as " Training Data Augmentation for Hidden Fruit Image Segmentation by using Deep Learning" [4], but has not been applied to semantic segmentation. Therefore, we propose and evaluate a data augmentation method using image composite for learning semantic segmentation.

In this study, we combine image composition with basic data augmentation methods such as cropping, flipping/rotation, and color manipulation to generate images suitable for learning multiple food ingredients. In the experiment, we checked the effectiveness of each processing.

2. Data augmentation method

2.1. *Cropping*

The purpose of this process is to generate various variations of food arrangement from a single image. A random area of 1000x1000 pixels is cropped from the input image.

2.2. *Horizonal Flipping and Rotation*

The purpose of this process is to increase the variation in the direction of rotation of the food. In the flipping process, the input image is flipped left to right. In the rotation process, the input image is rotated by affine transformation in the range of 0 to 360 degrees around the image center. In a series of processing, the input image is flipped with a probability of 50 percent, followed by rotation.

2.3. *Composition*

The purpose of this process is to deal with overlapping of foodstuffs. First, we cut out only the region of food parts of the input image. Next, we shrink the region so as not to introduce noise at the boundary with the background, and merge it into another input image. After merging, unnatural unevenness is created in the boundary area, so finally, blurring is applied only to the boundary area. Since the actual image to be recognized can be assumed to show only one type of food or all types of food, the compositing process is performed random times for one training image.

2.4. *Color Jittering*

The purpose of this process is to capture essential features such as patterns rather than the color of the food. Each of the three RGB values of the input image is multiplied by 0.8 to 1.2.

3. Experiments in food recognition

3.1. *Experiment Preparation*

Carrots, cabbage, sprouts, pork, green peppers, onions, and shimeji mushrooms were prepared and stir-fried over medium heat for 5 minutes before being photographed.

The shooting conditions were as follows: Panasonic DMC-FZ200 camera, f-number 8.0, SS 1.0, ISO sensitivity 100, distance between the camera and the plate 60cm, zoom 4x, under fluorescent light, and in forward light.

After cooking the food, we divided the food into two groups: one for images of one kind food and the other for images of mixed food. 28 images of one kind food were taken, 4 for each type of food, and 8 images of mixed food were taken, 4 for training data and 4 for test data. For the images of mixed food, 8 images were taken, 4 for training data and 4 for test data. The same food ingredients were not photographed more than once.

3.2. *Experimental method*

3.2.1. *Learning and Recognition*

In training, the model is trained on a training image of 1000x1000 pixels after data augmentation.

In recognition, test data is input to the trained model for recognition. The image input as test data is 4000x3000 pixels, which is different from the trained 1000x1000 pixels training image. Therefore, we split the input image, perform semantic segmentation on each of them, and then merge them into the original form to output the recognition result.

3.2.2. *Evaluation function*

As a metric for evaluating the performance of a multi-class classification mode, there is an evaluation function called categorical accuracy provided by the machine learning library keras [5]. The categorical accuracy is the number of pixels where the correct label matches the label with the highest output value of the model, divided by the total number of pixels. Here, since the images prepared in this study are easy to recognize the background. The more the background area of the test image is, the higher the categorical accuracy will be. Therefore, it may not be possible to properly evaluate the percentage of correct answers for the food ingredients. For this reason, we use our own evaluation function, food accuracy, which evaluates the correctness rate using the same method as categorical accuracy for only the pixels of foodstuff excluding the background.

3.3. Experimental conditions

The common training conditions are as follows
Model: U-Net [6], Number of training data: 10000, Number of epochs: 30, Batch size: 8, Loss function: Categorical Cross Entropy, Optimizer: Adam [7], Learning rate: 0.001.

As shown in Table **1**, The experiment was conducted in six different training conditions. Each condition is shown below.

- Training condition A
  Training with training data generated by color jittering, flipping/rotation, cropping.
- Training condition B
  Training with training data generated without color jittering in the condition of training A.
- Training condition C
  Training with training data generated without flipping/rotation in the condition of training A.
- Training condition D
  Training with training data generated without cropping in the condition of training A.
- Training condition E
  Training with training data generated without compositing in the condition of training A.
- Training condition F
  Training with training data generated without compositing using images of mixed food instead of images of one kind food in the condition of training A.

By comparing the results of training condition A with those of the other conditions, we can see the effects of each data augmentation process.

Table 1. Training conditions

Training	**Image type for training**	**Color jittering**	**Flipping and rotation**	**Cropping**	**Composition**
A	**One Kind**	○	○	○	○
B	**One Kind**		○	○	○
C	**One Kind**	○		○	○
D	**One Kind**	○	○		○
E	**One Kind**	○	○	○	
F	**Mixed**	○	○	○	-

Table 2. Food accuracy after Training (%)

Training	Food Accuracy
A	86.0
B	82.7
C	86.6
D	77.5
E	48.3
F	85.7

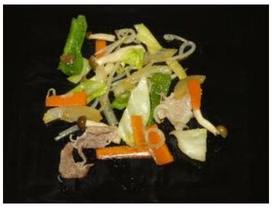


Fig. 1. Input Image

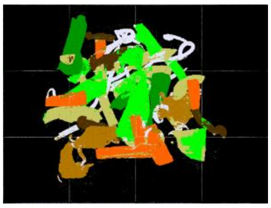


Fig. 2. Output Image

3.4. Experimental results

The food accuracy after training is shown in Table 2. An input image and the output image of the model of training A for it are shown in Fig. 1-Fig. 2.

4. Discussion of the experimental results

In Table 2, the food accuracy of training B, D, and E was lower than that of training A. This confirmed the increase of food accuracy by data augmentation of color jittering, cropping, and composition. On the other hand, there was almost no difference in food accuracy between training A and C. Therefore, we could not confirm the increase of food accuracy by data augmentation of inversion and rotation. This is because there was enough variation in the training data in terms of food orientation that flipping and rotation became meaningless data augmentation.

Since there was almost no difference in Food Accuracy between training A and F, it was confirmed that the combined image of images of one kind food can be used as training data equivalent to the image of mixed food.

5. Conclusions

We proposed the method for the semantic segmentation of food images allows the model to learn efficiently by composing images of one kind food that can be easily labeled. As for the data augmentation methods, we confirmed the increase of Food Accuracy by the data augmentation of color jittering, cropping, and composition.

References

1. Keiji Yanai. "Research Trends on Food Image Recognition." journal of the Japanese Society for Artificial Intelligence 34.1 (2019): 41-49.
2. Hirotaka Tanaka, and Hiroyuki Shinnou. "Efficient Construction of training data for Object Detection." The 34th Annual Conference of the Japanese Society for Artificial Intelligence. The Japanese Society for Artificial Intelligence, 2020.
3. Mikołajczyk, Agnieszka, and Michał Grochowski. "Data augmentation for improving deep learning in image classification problem." 2018 international interdisciplinary PhD workshop (IIPhDW). IEEE, 2018.
4. Ryoma Takai, and Kazuki Kobayashi. "Training Data Augmentation for Hidden Fruit Image Segmentation by using Deep Learning." The 33rd Annual Conference of the Japanese Society for Artificial Intelligence. The Japanese Society for Artificial Intelligence, 2019.
5. Francois Chollet, et al. Keras. https://keras.io, 2015.
6. Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.
7. Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).

Authors Introduction

Mr. Takayuki Yamabe

He received his B.S. degree in Engineering in 2021 from the School of Mechanical Engineering, Kanazawa University in Japan. He is acquiring the M.E. in Kanazawa University.

Mr. Tatsuya Ishichi

He is acquiring the B.S. degree in Kanazawa University.

Dr. Tokuo Tsuji

He is an associate professor of Institute of Science and Technology at Kanazawa University in Japan.

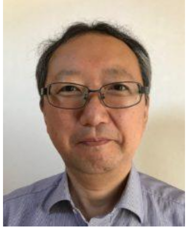
He received his D. Eng. degree from Kyushu University in 2005.

His research interests include manipulation, image processing, and robotic hand.

Dr. Tatsuhiro Hiramitsu

He is assistant professor of Institute of Science and Engineering, Kanazawa University in Japan. He received Dr E. degrees from school of engineering, Tokyo Institute of Technology, Japan in 2019. His research interest is in the soft structure mechanisms for robotic systems.

Dr. Hiroaki Seki



He is a Professor of Institute of Science and Technology at Kanazawa University in Japan. He received his D. Eng. degree in Precision Machinery Engineering from the University of Tokyo in 1996. His research interest includes novel mechanism and sensing in robotics and mechatronics.