4Q-06

CNN-BASED VISIBLE INGREDIENTS RECOGNITION IN A FOOD IMAGE USING DECISION MAKING SCHEMES

Kun Fu† Ying Dai† Ziyi Zhu†

Faculty of Software and Information Science, Iwate Prefectural University†

1. INTRODUCTION

In recent years, there are more and more patients with diabetes. One of the important reasons is that many people have unhealthy lifestyles and bad eating habits, Therefore, the study on food image recognition has very important practical significance.

In this paper, we focus on recognizing the ingredients segmented from the food images, and propose a method for it. In details, the candidate regions of the ingredients for the recognition are located by the methods of locating and sliding windows. Then, these regions are assigned to the ingredient classes by the CNN-based ingredients classifier trained by the dataset with single ingredient images. Finally, the ingredients are decided from these candidate results by using the scheme of decision making. The effectiveness of the proposed method is evaluated by the experimental results.

2. RELATED WORK

2.1. Transfer learning

In this paper, the transfer learning based on EfficientNetB0 and Resnet are performed to obtain the two models for the single ingredient recognition.

2.2. Decision making

Decision making is a method of making decisions in the case of disagreement. In this study, it was used to make the final decision when different results emerged during the component identification process.

2.3.CNN-based visible ingredient segmentation [1]

This study segmented all the visible ingredients in the food images. However, there are some problems with the segmentation results. Some of the ingredients are separated into different segments, and some of the different ingredients are returned as the same segmentation. Our paper uses the segmentation results of [1] to continue the ingredient recognition, with focusing on the above problems.

3. DATASET

3.1. Single ingredient dataset (SI dataset)

SI dataset contains 10750 ingredient images, involving in 107 kinks of ingredients. This dataset is used to train the single ingredient recognition model.

3.2 Multi-ingredient food image dataset (MIFI dataset)

MIFI dataset contains a total of 2166 images. Each image contains a multitude of ingredients. This dataset is used to evaluate the performance of the ingredient recognition in the food images.

4. METHOD

In this section, we propose a framework of multi-ingredient recognition system, which is shown in Figure 1. For this system, the candidate regions of the segments for the recognition are located by the methods of locating and sliding windows. Then, these regions are assigned to the ingredient classes by the CNN-based single ingredient recognition model trained by the above single ingredient dataset. The ingredients in the segments are finally decided from these candidate results by using the scheme of making decisions.

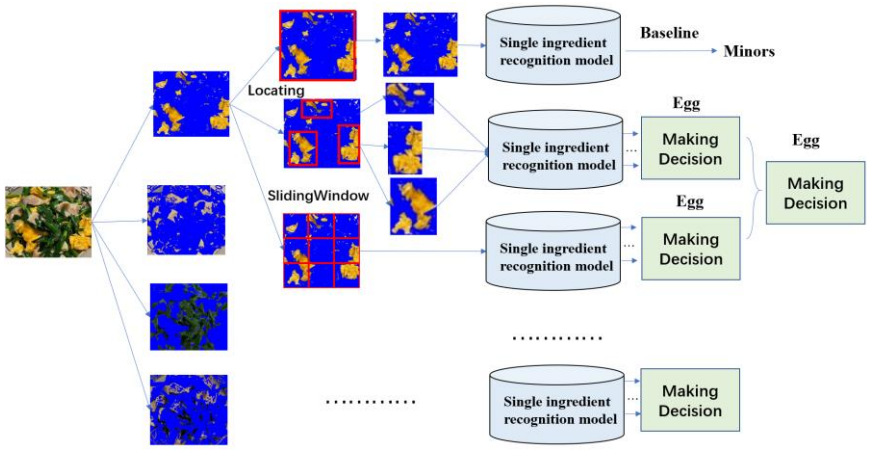

Fig 1. Multi-ingredient identification system

4.1. Feeding the segments into the model directly

That the segment is directly feed into the single ingredient recognition model without any processing is used as a baseline.

4.2. Locating candidate regions

Locating candidate regions of the ingredients is by first grayscale processing of the segmented image and then finding the ingredient regions in the image by erosion and dilation, Finally the boxed ingredient area is locked by measuring properties. An example is shown in Figure 2.

Fig 2. Locating the candidate regions

4.3. Sliding window

The sliding window intercepts only a portion of the image area at a time for ingredient recognition. Every sliding window is feed to the single ingredient recognition model for classification.

4.4. Making decision

We design two algorithms of making decisions for the ingredient identification, which combine the methods of locating candidate region and sliding window. Algorithm 1 is under the premise of one segment image having one ingredient. Algorithm 2 can return multiple ingredient recognition results for each segment image, considering the limitation of [1].

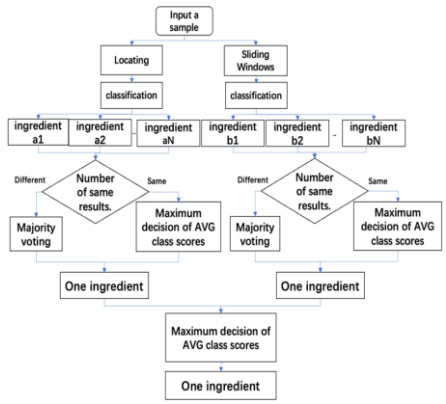

Fig 3. Algorithm 1 flowchart

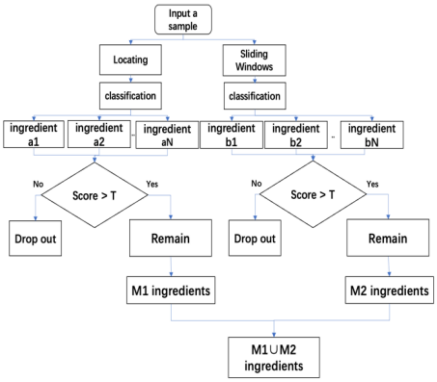

Fig 4. Algorithm 2 flowchart

5. EXPERIMENTS AND ALALYSIS

We use the precision, recall, and F1-score for evaluating the performance of the ingredient recognition in the food images.

5.1. Experimental results and analysis

The results in Table 1 show that directly feeding the segmented images into the model without any processing (Baseline) does not give the good results. On the Efficientnet, the median of average F1 score for Algorithm 2 reaches 37.5%.

Table 1. Experimental results of Baseline, Algorithm 1 and Algorithm 2.

Method	Model		AVG Precision	AVG Recall	AVG F1score
BaseLine	Effcientnet	median	33.33	16.74	20.00
	Resnet	median	30.56	11.81	16.11
Algorithm1	Effcientnet	median	55.88	32.13	35.90
	Resnet	median	45.02	30.52	33.03
Algorithm2	Effcientnet	median	39.84	49.91	37.50
	Resnet	median	42.73	37.03	35.09

From the experimental results in Table 2, we can see that compared to Algorithm 2, for the segments, the medians of the average F1score of the sliding-only drops about 10%, and those of locating-only drops about 17%. And in order to verify that the ingredient segmentation is helpful for recognition, we have an investigation of sliding windows on the original image directly. The experimental result shows that the medians of the average F1score drops about 10% compared to the results of Algorithm 2.

Table 2. Results of sliding alone or locating alone

Method	Model		AVG Precision	AVG Recall	AVG F1score
Original image sliding-only	Effcientnet	median	32.54	48.55	26.36
	Resnet	median	30.32	42.25	23.42
Segmented image sliding-only	Effcientnet	median	25.00	58.55	27.46
	Resnet	median	23.00	53.24	24.03
Segmented image locating-only	Effcientnet	median	54.04	10.49	19.75
	Resnet	median	52.04	10.02	18.00

6. CONCLUSION

In this paper, we proposed a new framework to recognize the ingredients in the food images. The experimental results verified the effectiveness of combining the locating method with the sliding window method.

7. REFERENCES

[1] Ziyi Zhu, Ying Dai, “CNN-based visible ingredient segmentation in food images for food ingredient recognition”, Proc. of AAAI AAI 2022, pp. 348-253, 2022, Japan.