A Fragmented Target Recognition System Based on Zero-Shot Learning

Chenqing Ji, Yujie Lu, Yongjuan Shi, Guang Wu*
wug@sustech.edu.cn
Department of Electrical and Electronic Engineering
Southern University of Science and Technology
Shenzhen, China

2023 IEEE International Conference on Consumer Electronics (ICCE) | 978-1-6654-9130-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCE56470.2023.10043466

***Abstract*—In recent years, target recognition and detection methods based on deep learning have shown great application prospects in many fields, such as smart house, driverless technology, product detection and military equipment, etc. However, in some extreme application scenarios, such as the emergency rescue, the target is inevitably fragmented due to the impact of explosion and many other factors, which leads the lack of effective feature information in the target image and affects the accuracy of target recognition and classification. In order to solve this problem, this paper proposes a new method to recognize fragmented targets based on zero-shot learning. This method solves the problem of target recognition under the condition of zero samples by introducing some high-level attributes. For verifying the effectiveness of this method, this paper takes five kinds of ingredients after cutting in daily life: cucumber, potato, tomato, eggplant, and bamboo as an example to illustrate and verify the whole process of the feature extraction, attribute recognition and the target classification in this method. The experiment results show that the highest recognition accuracy for the ingredients after processing in this fragmented recognition system is 76%. In addition, this paper also develops and verifies the real-time recognition of this system on the embedded platform PYNQ.**

Keywords—Zero-Shot Recognition Algorithm, Ingredients recognition, Neural Network, PYNQ

I. Introduction

In recent years, deep learning-based target recognition and detection methods have shown great promise in many fields such as smart homes, unmanned vehicles, product inspection and military equipment. A method of recognition and classification of vegetable image based on deep learning, using the open-source deep learning framework of Caffe and the improved VGG network model to train the vegetable image data set and the accuracy rate was as high as 96.5% and in VGG network it was 92.1% [1].

However, in some extreme application scenarios such as the emergency rescue scenarios, targets are inevitably fragmented due to some unpredictable factors, such as collision, explosion, or collapse. Therefore, the missing effective feature information of the target image will dramatically decrease the accuracy of target recognition or classification. In this paper, fragmented target means the target splits into several parts. As shown in the Fig. 1, the first line shows the entire tomatoes, cucumbers, potatoes and the second line shows the fragmented tomatoes, cucumbers and potatoes respectively.

Generally speaking, the images of fragmented targets are difficult to collect, resulting in a small number of samples, which makes the traditional deep learning methods almost fail and further increases the difficulty of the fragmented target recognition. However, the zero-shot learning model provides an effective idea for solving this recognition problem under the condition of less target samples. The training and testing classes in zero-shot learning model are mutually exclusive and need to be completed by knowledge transfer between the training and testing classes. Therefore, zero-shot learning is also a special scenario of transfer learning [2].

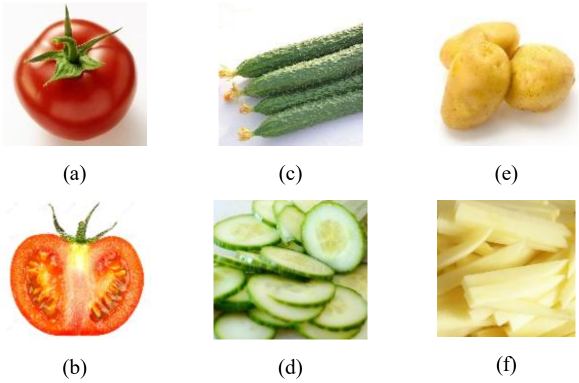

Fig. 1. Intact state and fragmented state of ingredients

(a) Intact tomatoes, (b) Fragmented tomatoes, (c) Intact cucumbers, (d) Fragmented cucumbers, (e) Intact potatoes, (f) Fragmented potatoes,

With the large-scale applications of convolutional neural network, Lili Pan et al. proposed a convolutional neural network recognition method based on the automatic multi-class classification, which explored three features such as PCA (Principal Component Analysis), CFS (Correlation-based Feature Selection), and IG (Information Gain). The recognition effect of the evaluator was as high as 87.78% under the ResNet network [3]. Further, in 2019, Lili Pan et al. proposed an innovative recognition method by using image transformation technology to expand a small ingredient data set, using transfer learning to extract image features and deep feature vectors to identify ingredient category [4]. However, the test images by this method were also intact ingredients.

In 2009, Lampert et al. proposed the DAP (Direct Attribute Prediction) model, which was the first model of zero-shot learning applied to the field of computer vision [5]. The model trains a classifier for each attribute of the input in the training phase and then the resulting model is used to predict the attribute. In the testing phase, the attributes of the test samples are predicted and the closest class from the vector space is found, which is the final recognition result.

In this paper, we conduct a recognition research on cut or processed ingredients by zero-shot learning algorithm. Zero-shot learning means learning a problem in which no training data are available for some classes or tasks and only the descriptions of these classes are given [6]. Compared with

978-1-6654-9130-3/23/$31.00 ©2023 IEEE

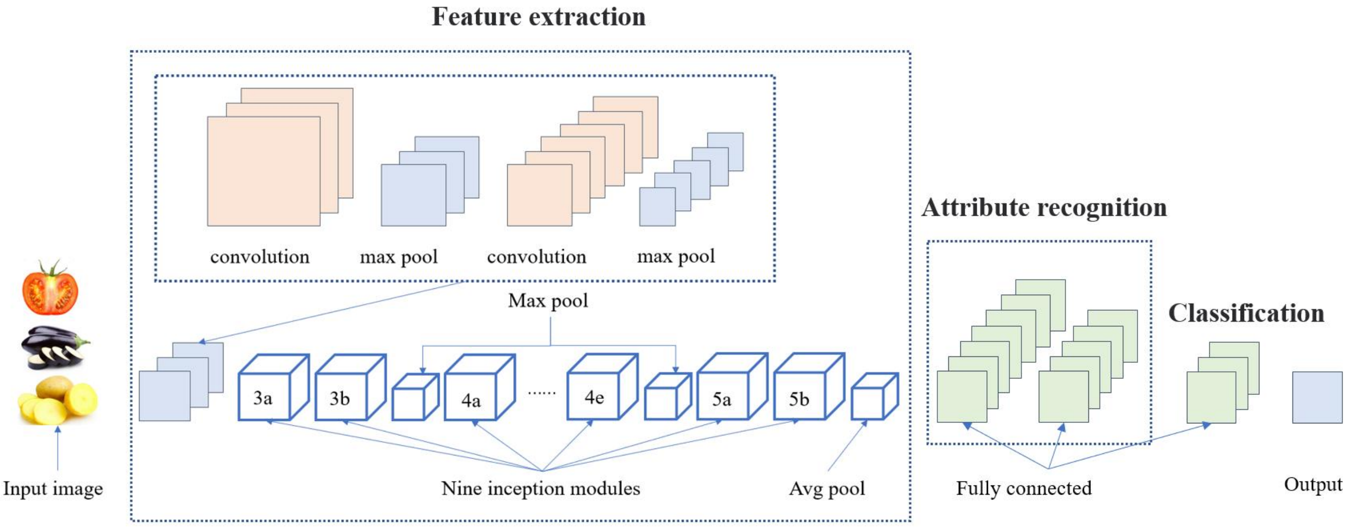

Fig. 2. The complete construction process of DAP model

traditional machine learning methods, the proposed system has the lower training cost, higher accuracy, and wider applications. What's more, the system can be used for real-time recognition of complex fragmented categories with few or no samples. The results show that in scenarios without training samples, our recognition system can perform intelligent ingredient recognition after the user inputs related high-level attributes features corresponding to the recognition target. In addition, this paper also realizes the real-time recognition on the embedded system PYNQ, which has advantages of real-time and low-cost.

In order to verify the effectiveness of the recognition method, we create 500 typical ingredient images data set named as FoodDataset5. The categories in this data set are cucumber, potato, tomato, eggplant, and bamboo. For each category, there exists 100 images, all of which are processed ingredients shape such as filaments, blocks, and flakes. At the same time, this paper also manually defines five high-level binary attributes, which are red, yellow, green, purple, and skinned respectively.

Moreover, this paper carries out the real-time recognition on x86 computer and embedded platform PYNQ respectively. For real-time recognition on x86 computer, we firstly train the whole model on software, then we perform the feature extraction, attribute recognition and classification model for each image stored on x86 computer and finally get the recognition result. Result shows that the highest recognition accuracy can reach 76% for the target ingredient data set and the recognition result on PYNQ is basically consistent with that on x86 computer.

II. Theoretical analysis of the algorithm

The DAP (Direct Attribute Prediction) model was proposed by Lampert et al. in 2009, which was the first model applied in the field of computer vision by zero-shot learning. The model builds a system to detect objects based on a list of human-specified high-level descriptions consisting of arbitrary semantic attributes, like shape, color, or even geographic information [5]. With these set of high-level attributes that act as an intermediate layer in the classification cascade, the zero-shot object detection can be achieved.

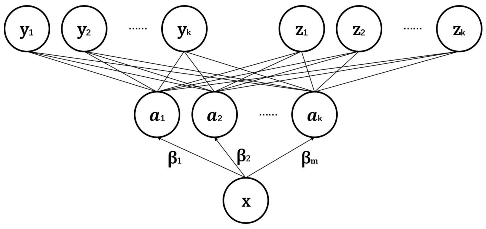

Fig. 3. The DAP (Direct Attribute Prediction) model

For each attribute a_{m}, there exists a classifier $\mathbf{f}_{a_{\mathrm{m}}}$ correspondingly. The probability estimates for each attribute of the input image are:

$$p(a|x) = \prod_{m=1}^{M} p(a_m|x) \tag{1}$$

In the testing process, the relationship between each attribute and zero-shot category is:

$$p(a|x) = [[a = a^z]] \tag{2}$$

When $a = a^z$, $\mathrm{p}(a|\mathrm{z}) = 1$. When $a! = a^z$, $\mathrm{p}(a|\mathrm{z}) = 0$. Applying the Bayesian formula, we get the relation of the attribute-category layer:

$$p(z|a) = \frac{p(z)}{p(a^z)}[[a = a^z]] \tag{3}$$

$$p(z|x) = \sum_{a\in\{0,1\}^M} p(z|a)p(a|x)$$
$$= \frac{p(z)}{p(a^z)} \prod_{m=1}^{M} p(a_m^z|x) \tag{4}$$

After calculating the posterior probability of each category, we obtain the final result of category recognition by mean precision prediction [5]:

$$f(x) = argmax \prod_{m=1}^{M} \frac{p(a_m^{z1}|x)}{p(a_m^{z1})}, l = 1, ..., L \tag{5}$$

The training data set and testing data set of the traditional CNN (Convolutional Neural Network) are intersected, which makes the neural network trained by training data set has poor transferability. Fig. 2 shows the complete construction process of the DAP model, we can see that the DAP model introduces a layer of self-defined attribute layers, which greatly enhance the transfer learning ability of convolutional neural network. The input layer of the neural network in DAP model is the pixel value of the image and the GoogLeNet convolutional neural network is used as a feature extractor to extract the feature information of an image for better perform the classification in the subsequent neural network.

III. System Building and Model Testing

The hardware of the ingredient recognition system is composed of camera module, processor module PYNQ and x86 computer. When applying the recognition system, the x86 computer receives the image data collected from the camera module and extracts the features of each image frame. Then, the processor module PYNQ receives the image features transmitted by computer and makes corresponding processing for each image feature. Finally, it displays the output results on x86 computer. This paper divides the whole recognition system into three sub-models: feature extraction model, attribute recognition model and classification model. These three sub-models are trained based on our self-made ingredient data set FoodDataset5. The design methods of these models are shown in the three sections below.

A. Feature extraction model

The function of feature extraction model is to extract the features from input images through multi-layer convolutional neural network for better attribute recognition. In this paper, GoogLeNet network, a typical feature extraction network, is selected to achieve feature extraction of the input images. Since this network is trained based on a large number of data sets, it can be viewed as an excellent feature extractor for the input ingredient images. In this paper, the output 1024 1×1 feature maps from the penultimate layer (Avg Pool layer) of the GoogLeNet network are selected as the final extracted feature tensor of each input image [7].

B. Attribute recognition model

The function of attribute recognition model is to map the extracted image features to the self-defined high-level binary attribute vectors. In this model, the input data is the feature vector at the size of 1×1024. Through two full connection layers, the output of this model is the corresponding high-level binary attribute vector. According to the selected training data set, this paper manually defines nine high-level binary attributes and each attribute can be reflected by two or more categories. For the value of these nine high-level binary attributes: 1 means that this attribute exists, while 0 means that this attribute does not exist. The codes of these nine high-level binary attributes and their corresponding explanations are shown in Table 1.

Table 1. Nine self-defined high-level binary attributes

code	**S**	**H**	**G**	**P**	**R**
Explanation	skinned	homogeneous	green	purple	red
code	**Y**	**F**	**C**	**W**	
Explanation	yellow	filamentous	circular	wedge	

In order to construct the attribute recognition model, we randomly disrupt the combinations of the above nine self-defined high-level binary attributes to determine which combinations of high-level binary attributes can make the recognition accuracy of the whole recognition system reach the highest. In the actual test, we select the fifth category of the ingredient data set--bamboo as the zero-shot target to recognize all the images of bamboo category (See part four for more specific description). The different combinations of the high-level binary attributes and their corresponding highest recognition accuracies are shown in Table 2 below.

Table 2. The recognition accuracy for zero-shot targets of this system by using different combinations of advanced binary attributes

High-level binary attributes composition (listed in their code)	The highest recognition accuracy
C F W H S	4%
R C F H S	3%
R Y F H S	6%
R Y G F S	12%
R Y G F H	10%
R Y G H S	3%
R Y G P S	76%
R Y G P H	6%
R Y G P F	71%
R Y G P C	68%
R Y G P W	69%
R Y G P S H	4%
R Y G P S F	46%
R Y G P S C	42%
R Y G P S W	38%
R Y G P S H F C W	2%

It can be seen from the above results that when using the high-level binary attributes combination red, yellow, green, purple, and skinned to build the high-level binary attribute vector of the attribute recognition model, the recognition system has the highest accuracy in recognizing zero-shot targets. At the same time, when the number of high-level attributes in the combinations of high-level binary attribute is five, with the increase of the number of color attributes, the recognition accuracy basically shows an upward trend. This is because the essence of the feature extraction model is to obtain the feature information of the three colors of red, green, and blue (RGB) in the ingredient images. Naturally, the greater the number of color attributes in the high-level binary attribute vector, the better it can fit the color features extracted by the neural network in the previous layer model, thus making the recognition accuracy of the system higher. Moreover, when the number of color attributes in the combinations of high-level binary attribute remains constant, the recognition accuracy of the system reaches the highest when the number of high-level attributes is five. Then, if the number of high-level attributes continues to increase, the recognition accuracy of the system will decline. This is because that the neural network connection in this case is too complex, resulting in the mappings from feature to attribute and attribute to category become relatively fuzzy. Based on the above analysis, we choose the high-level binary attributes combination red, yellow, green, purple and skinned (The corresponding attribute code is <R,Y,G,P,S>) to build the attribute recognition model.

C. Classification model

The function of the classification model is to output the corresponding classification results of the input high-level binary attribute vector with the size of 1×5. The input of this model is the binary attribute vector with the size of 1×5, and the output of this model is the corresponding tensor of five categories. The network in this model adopts a simple full connection layer to establish the relationship from attributes to categories.

The testing process of the whole model is shown in Fig. 4. Firstly, we read the images by external USB camera and then process these images by the three models mentioned above. After that, we can obtain the classification result at the output of the classification model. The application on the embedded system PYNQ is similar to that on x86 computer and the slight difference is that during the test, the feature tensors of image data are still firstly extracted by x86 computer. Then, they are uploaded to PYNQ by SFTP protocol. Finally, the ARM on PYNQ completes the whole calculation process of attribute recognition model, classification model and displays the recognition results on x86 computer.

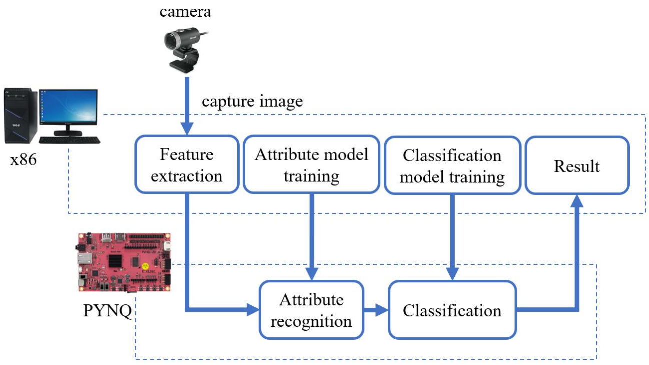

Fig. 4. Flow chart of the whole system

IV. MODEL TESTING AND APPLICATION

This part mainly introduces the realization of the ingredient recognition system on x86 computer and the embedded system PYNQ. Section A is a brief introduction to show how to train whole models of ingredient recognition system correctly. Section B introduces the testing result for zero-shot recognition of each target ingredient on the x86 computer. Section C compares the working performance of our model with the existing recognition schemes. Section D introduces the application of the system on the embedded system PYNQ.

A. Model training

1) Training attribute recognition model

In the training of the attribute recognition model, we use the images in our ingredient data set FoodDataset5 to train this model. Firstly, we select a certain category of fragmented ingredients and make it not be trained. Then, all the other categories of ingredients in our data set are trained, which creates the condition for zero-shot recognition. During the training process, five self-defined high-level binary attributes are needed to be iterated repeatedly, for which the model can learn the mapping relationship from the feature maps to five high-level binary attributes accurately. The training result shows that the recognition accuracy of these five attributes fluctuate at the early stage of training. As the number of the iteration epochs increases, the recognition accuracy of these five attributes shows an increasing trend and reaches a stable state when the number of iterations is about 50.

2) Training classification model

In the training of the classification model, we define five high-level binary attributes and the corresponding category for each ingredient image in order to train the model to learn the mapping relationship between them. Through the training process, the highest classification accuracy of this model can reach 98% and the classification accuracy reaches a stable level when the number of iteration epochs is about 1/6 of the total number of iteration epochs.

B. Model testing

After all the sub-models are trained, we test the whole system by using a category of fragmented ingredient images which have not been used in training process on the x86 computer. Before model testing, we select a certain category of fragmented ingredient and let its images not be trained in the attribute recognition model but be trained in the classification model. During the test, the images belong to this certain ingredient category are firstly processed by the feature extraction model to obtain the feature vector with the size of 1×1024 as the input of attribute recognition model. After through the attribute recognition model, the binary attribute vector with the size of 1×5 is obtained and it also be used as the input of the classification model. Finally, through the classification model, the system outputs the category tensor of this certain category of ingredient on x86 computer.

During the actual test, we select total five categories of ingredients in our ingredient data set FoodDataset5 and train the model mentioned in section A for each fragmented ingredient category to build the conditions for zero-shot recognition respectively. After that, we selected 100 images of each ingredient category as the testing data set to conduct the recognition test. The recognition accuracy in this test is presented by the confusion matrix in Fig. 5.

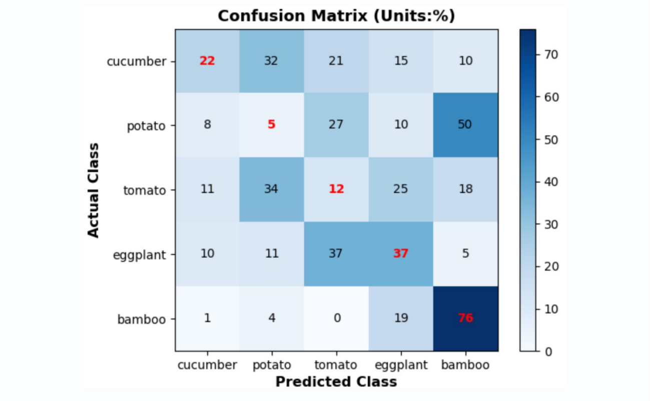

Fig.5. The confusion matrix in this section (The text highlighted in red is the correct recognition accuracy for each category of ingredient)

From Fig. 5, for the five categories of fragmented ingredients, only bamboo achieves the highest recognition accuracy of 76%, while the highest recognition accuracies of the other four categories of ingredients are all lower than 50%. Even worse, the recognition accuracy of potato and tomato are only 5% and 12% respectively. By analyzing the Fig. 5, the probability of recognizing potato as bamboo reaches 50%. It can be seen in the potato data set that many potato images are very close to bamboo in color, which can cause great interference for the model to recognize potato without any potato samples. Based on that, the model will recognize most of the potato images as bamboo. For tomato, the probability of the model recognizes it as potato and eggplant are 34% and 25% respectively. It is due to the selection of potato and eggplant data sets that many images contain red element, making the model falsely recognizes the red tomato as potato or eggplant. For cucumber, the probability of model classifies it into these five categories of ingredients is more average than that of other four ingredients categories. This is because the definition of the "green" attribute is not accurate. Therefore, the probability of the model falsely recognizes cucumber as the other four categories of ingredients is more average. For eggplant, the probability of the model falsely recognizes it as tomato and correctly recognizes it are both 37%, which is due to the fact that red and purple are the adjacent color among the five high-level binary attributes. Therefore, when the system recognizes eggplant with zero samples, it is easy to misjudge the "purple" attribute as red and then falsely recognizes it as tomato.

In summary, when the DAP model is used for zero-shot recognition for five categories of fragmented ingredients, the highest recognition accuracy is 76% when recognizing the bamboo. However, due to the relative simplicity of the five self-defined high-level binary attributes and the interference of many mixed ingredients or minor color in our selected data set, our model becomes relatively poor in recognizing a certain of ingredients.

C. Model comparison

In order to further illustrate the feasibility of establishing DAP model to achieve zero-shot recognition of the target ingredients, we also compare this model with several other feasible ingredients recognition schemes. The model based on the compared system uses MobileNet_v1 as the feature extraction neural network and uses Single Shot MultiBox Detector (SSD) as the classifier. The core idea of MobileNet is to introduce depthwise separable convolution, which splits the standard convolution filter into two structures: depthwise convolution and pointwise convolution. SSD performs object detection work on feature maps of multiple scales, so that targets of all scales can be taken into account: small-scale feature maps predict large targets, and large-scale feature maps predict relatively small targets [8].

We will further illustrate the feasibility of the DAP algorithm in the field of fragmented ingredients recognition through five following comparative tests results in Table 4.

Table 4. The recognition accuracy in this subsection

Ingredient	bamboo	cucumber	potato	tomato	eggplant
Test 1	79%	87%	88%	96%	83%
Test 2	5%	7%	11%	8%	3%
Test 3	67%	57%	52%	74%	63%
Test 4	0%	0%	0%	0%	0%
Test 5	94%	81%	83%	90%	81%

In test 1, there are five complete ingredients categories appear in both training set and testing set. In the training set, there are 32 fragmented images for each of the five ingredients categories. In the test set, there are 8 fragmented images for each of the five ingredients categories. Then, we set parameters such as batch size, class numbers, and step numbers in the configuration file according to the labeling situation and training objectives. After tuning, the batch size, class number and the step number are set to 8, 4 and 80000 respectively. Finally, the recognition accuracy of various ingredients categories is shown in the first line of Table 4. In test 2, there are five complete ingredients categories in the training set and five fragmented ingredients categories in the testing set. In test 3, there are five fragmented ingredients categories appearing in both training and testing set. In test 4, there are four completed ingredients categories in training set and one new completed ingredient category in testing set.

From Table 4, using the traditional object recognition network, the complete and fragmented ingredients images can be more accurately recognized. If the training set contains five complete ingredients categories but the testing set contains the corresponding five fragmented ingredients categories, we can find in test 2 that the recognition rate is low and almost less than 10%. However, if the training set is four complete ingredients categories but the testing set is another new complete ingredient category, result in test 4 shows that the traditional convolutional neural network cannot recognize the new ingredient category.

In order to form a blank contrast with the basic test of the model in Section B, we let the images of five categories of ingredients appear in the training set and the testing set and use the same test method in Section B to recognize these ingredient images on x86 computer. Based on that, we calculate the recognition accuracy of the system shown in test 5 in Table 4. Compared with the test result in Section B, when all categories of ingredients appear in training set, the recognition accuracy of the model is further improved and the recognition accuracy of bamboo is increased from 76% in Section B to 94%. However, although the accuracy of the model in recognizing bamboo has been improved, if the bamboo data set is relatively rare, it is not worthwhile to spend a lot of time and resources to find the data set of rare ingredients in the pursuit of higher recognition accuracy. Therefore, using DAP model to recognize the target which has zero samples has great advantages.

D. Model application

In order to further apply the results of the previous section to actual ingredients recognition, we take the images of a certain ingredient category by USB camera and try to calculate which one of the five defined ingredients categories belongs to this image by our ingredient recognition system on the embedded development board PYNQ. The use of the PYNQ can reduce the computational energy consumption sufficiently to recognize ingredients efficiently compared to x86 computer. This innovative application idea can enable us to achieve low-power recognition of the target ingredients.

In this Section, we connect USB camera to x86 computer to take images of different ingredients. In order to enhance the interaction of this recognition system, the customer can select whether to recognize an image by pressing a specific key by our design. When the specified key is pressed, the image obtained will firstly enter into the feature extraction model on the x86 computer for the deeper image feature extraction. After that, the image features extracted by x86 computer are transmitted to the ARM on PYNQ in real time through the SFTP protocol. Then, these features are processed by attribute recognition model and finally we obtain the recognition results by classification model on PYNQ.

For the five categories of ingredients shown in the above section, this paper conducts two tests: In test 1, for each category of ingredients, we select a typical image that reflects this ingredient category and repeat 50 consecutive recognition tests. In test 2, for each category of ingredients, we select 30 images in the data set of each ingredient randomly and each image is tested for three times. If at least one time can be recognized accurately, we can say that our system can recognize this image. The recognition result for each test is shown in Table 5. Meanwhile, in order to demonstrate the reliability of these two tests, we also conduct them on x86 computer. The recognition accuracy of each test on x86 computer compared with that on embedded system PYNQ is shown in Fig. 6 below.

Table 5. The recognition accuracy for two tests on PYNQ

Ingredient	bamboo	cucumber	potato	tomato	eggplant
Test 1	80%	8%	72%	100%	64%
Test 2	80%	12%	13.3%	86.7%	66.7%

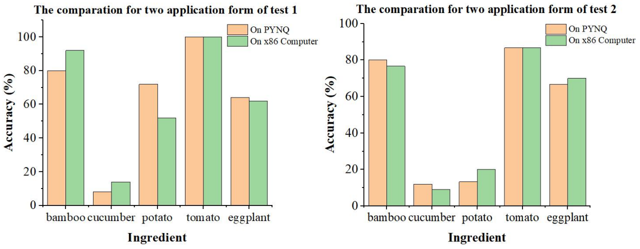

Fig. 6. The comparation for two application form of test 1 and test 2

In test 1, for tomato, the recognition accuracy is up to 100%, which reflects that red attribute is relatively easy to learn in attribute recognition model. The bamboo can be recognized with the recognition accuracy of 80%, which may due to the good selection of the testing images in bamboo data set. The recognition accuracy of potato and eggplant is around 65%, indicating that the model could not correctly classify the images of potato or eggplant in each recognition. However, for cucumber, the recognition accuracy is only 8%, which is because there exist many images of the mixed ingredients in the cucumber data set and these images cause great interference to the learning of "green" attribute by our system. Therefore, the recognition accuracy of cucumber is very low.

Compared with test 1, test 2 selects 30 different pictures for each category of ingredients, so the test result can better demonstrate the universality of the recognition system. In test 2, the recognition accuracy of tomato, bamboo and eggplant can reach more than 65%, and the average accuracy of tomato is up to 86.7%. However, for potato and cucumber, the recognition accuracy is only 13.3% and 12% respectively. The main reason for the low accuracy of these two ingredients categories is that: For potato, the color of some images randomly selected in potato data set is very close to bamboo, so these images can easily be recognized as bamboo when taken by camera, which resulted in a much lower recognition accuracy of potato than test 1. For cucumber, in addition to the reasons mentioned in the analysis of test 1, as many images in cucumber data set are peeled, the color of the surface after peeling is quite different from the color of the cucumber skin, which will cause great interference to the learning of "green" attribute by our model. Therefore, in test 2, our system still has the worst recognition accuracy on cucumber.

From the above analysis, we can conclude that when the recognition system is applied on embedded system PYNQ, the recognition accuracy of tomato and bamboo is the best and the recognition robustness and universality are high as well. For potato, although the robustness of the system to recognize the same image is not poor, but the universality of the system in the recognition of multiple images is poor. However, for cucumber, the robustness and universality of this system is the worst. At the same time, by comparing the results of the same recognition tests on PYNQ and x86 computer, it can be seen from Fig. 6 that in both tests, the recognition accuracy of the five categories of ingredients on PYNQ and x86 computer are similar, which reflects the reliability of our model application on embedded system PYNQ.

Acknowledgment

The authors would like to acknowledge the financial support from the Special Funds for the Cultivation of Guangdong College Students' Scientific and Technological Innovation (Grant No. pdjh2022c0024).

V. Conclusion and Prospects

This paper innovatively proposes a fragmented target recognition algorithm based on GoogLeNet network and implements it in real time on PYNQ. In the composition of the whole recognition system, GoogLeNet network is used for feature extraction, attribute recognition model is used to establish the relationship between the features of input images and five high-level binary attributes. The highest recognition accuracy of classification model can achieve 76% in testing the untrained fragmented ingredient on x86 computer. The test results in this paper can be applied to the recognition of fragmented ingredients with small or zero samples, which is widely used in real life. In the future, we will focus on optimizing the self-made ingredient data set and high-level binary attributes so as to accurately recognize each single category of ingredients when facing the mixed ingredients. At the same time, we will also try to expand the self-made ingredient data set FoodDataset5 and change the ingredient categories in the training and testing data sets so that our recognition system can be widely applied to the recognition of more ingredients categories with only small or zero samples.

References

[1] Li, Z., Li, F., Zhu, L., & Yue, J. (2020). Vegetable Recognition and Classification Based on Improved VGG Deep Learning Network Model. *Int. J. Comput. Intell. Syst., 13*, 559-564.

[2] Wei Wang, Vincent W. Zheng, Han Yu, and Chunyan Miao. 2019. A Survey of Zero-Shot Learning: Settings, Methods, and Applications. ACM Trans. Intell. Syst. Technol. 10, 2, Article 13 (March 2019), 37 pages. https://doi.org/10.1145/3293318

[3] L. Pan, S. Pouyanfar, H. Chen, J. Qin and S. -C. Chen, "DeepFood: Automatic Multi-Class Classification of Food Ingredients Using Deep Learning," 2017 IEEE 3rd International Conference on Collaboration and Internet Computing (CIC), 2017, pp. 181-189, doi: 10.1109/CIC.2017.00033.

[4] Pan, L., Qin, J., Chen, H., Xiang, X., Li, C., & Chen, R. (2019). Image augmentation-based food recognition with convolutional neural networks. Computers, Materials, & Continua, 59(1), 297-313. doi:http://dx.doi.org/10.32604/cmc.2019.04097

[5] C. H. Lampert, H. Nickisch and S. Harmeling, "Learning to detect unseen object classes by between-class attribute transfer," 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 951-958, doi: 10.1109/CVPR.2009.5206594.

[6] Larochelle, H. , Erhan, D. , & Bengio, A. Y. . (2008). Zero-data Learning of New Tasks.

[7] C. Szegedy et al., "Going deeper with convolutions," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1-9, doi: 10.1109/CVPR.2015.7298594.

[8] Howard, A. G. , Zhu, M. , Chen, B. , Kalenichenko, D. , Wang, W. , & Weyand, T. , et al. (2017). Mobilenets: efficient convolutional neural networks for mobile vision applications