Embedded Selforganizing Systems

Design and Implementation of Intelligent Vegetable Recognition System based on MobileNet

Linlin Yang
Heilongjiang Province Key Laboratory of Laser Spectroscopy Technology and Application
Harbin University of Science and Technology
Harbin, China
2120610125@stu.hrbust.edu.cn

Yingluo Song
Heilongjiang Province Key Laboratory of Laser Spectroscopy Technology and Application,
Harbin University of Science and Technology
Harbin, China
2120610140@stu.hrbust.edu.cn

Zilin Hu
City University of Hong Kong
Hongkong, China
hzlinininin@163.com

Aili Wang*
Heilongjiang Province Key Laboratory of Laser Spectroscopy Technology and Application
Harbin University of Science and Technology
Harbin, China
aili925@hrbust.edu.cn

Haibin Wu
Heilongjiang Province Key Laboratory of Laser Spectroscopy Technology and Application
Harbin University of Science and Technology
Harbin, China
woo@hrbust.edu.cn

Yuji Iwahori
Department of Computer Science
Chubu University
Aichi, Japan
iwahori@isc.chubu.ac.jp

***Abstract*—With the development of food safety traceability and the rise of self-service supermarket. The automatic identification technology of agricultural products such as vegetables in circulation and sale has become an urgent problem. In this paper, we design an intelligent vegetable recognition system based on MobileNet, which includes the main control core, visual processing and other modules. The visual processing model is composed of Depth Separatable Convolution (DSC), which separates channels from regions and is therefore computationally efficient and suitable for embedded devices with low storage space. The results show that the overall recognition accuracy of the system for five vegetables is 97.33%, which has the advantages of stability, intelligence and convenience.**

Keywords—intelligent vegetable recognition, MobileNet, DSC, Embedded device.

I. INTRODUCTION

At present, vegetable sales mostly adopt the method of manual identification, which has many problems such as low efficiency of information collection, nonstandard classification, difficult safety traceability and high labor cost. The main occasions for vegetable circulation are retail stores, vegetable wholesale markets and large supermarkets.

Retail stores and vegetable wholesale markets input the unit price of various vegetables in electronic platform scales, electronic scales and other equipment for weighing and billing. Large supermarkets need staff to input PLU (price look up) code on the electronic scale to weigh and charge vegetables. Moreover, there may be input errors, which will delay a lot of time. When the number of customers is increasing, the queue will become longer and longer, which is not conducive to business sales and greatly reduces the shopping experience of customers. These two methods of selling vegetables are traditional and time-consuming, with a low degree of automation.

Most of the vegetable recognition methods based on traditional image processing need to manually define the color, texture and other features of vegetables, because the gray level co-occurrence matrix describing texture features takes up a large space and takes a long time to calculate; Classifiers (minimum nearest neighbor, support vector machines, decision trees, neural networks, etc.) need to extract a large number of features to achieve good training results, which has great constraints.

In order to realize the full automation of vegetable circulation, many researchers have applied image recognition technology to the field of vegetable classification. Arivazhagan et al. [1]. used color and texture to identify fruits, and verified the effectiveness in 15 kinds of fruits. Rocha et al. [2] combined different features and classifiers to reduce the classification error to 15% in 15 categories of fruits and vegetables with less training data. Dubey et al. [3] proposed an improved sum difference histogram texture feature description method, which was verified in the fruit and vegetable library. Dubey et al. [4] proposed a fruit classification method, using K-means for image segmentation and support vector machine(SVM) for classification. Experiments show that this method can effectively classify fruits. Sachin C et al. [5] applied YOLO algorithm to the field of vegetable recognition, input several different vegetable images into the network, manually draw the boundary box around the vegetables using OpenCV, and preprocess the images before training. The recognition accuracy of this method reached 61.6%.

MobileNet is an efficient and compact network proposed to solve the computing power constraint problem of small technical equipment. It is connected in series by deep convolution of basic units and separable convolution. This new convolution method can effectively extract the spatial and channel features of data by equivalent the standard convolution operation to two newly defined convolution operations.

MobileNet is a lightweight neural network, which is widely used in many fields by researchers. T Ghosh et al. [6] applied MobileNet to Bengal handwritten character recognition, which achieved 96.46% accuracy in recognizing 231 categories (171 compound characters, 50 basic characters and 10 numbers). Lei Y et al. [7] studied a human ear image recognition method based on SSD-MobileNet. The experimental results show that the recognition accuracy reaches more than 99%, and it has good robustness to images with background interference. Eko Prasetyo et al. [8] used MobileNet to classify the freshness of fish eyes, and used extended DSC to extract features and achieved good recognition results.

In this paper, MobileNet is applied to the field of vegetable recognition, and DSC is used to extract the color and texture features of vegetable images. Using the limited memory space of the embedded device to complete the classification task, it also has high recognition accuracy.

II. Design of Intelligent Identification System

In the system architecture of the intelligent vegetable recognition system based on MobileNet. STM32 chip is used as the main control core. Kendryte k210 Dock is used as the visual processing module, which communicates with MCU through serial port. HX711 pressure sensor is used as weighing module. HLK-V40 is used for text to voice broadcasting. OLED is responsible for display. When the pressure sensor detects that there are vegetables to be weighed, the visual processing module completes the vegetable classification, broadcasts the vegetable's name, unit price and total price by voice, and displays the weight, unit price and total price by OLED. The system architecture is shown in Fig. 1.

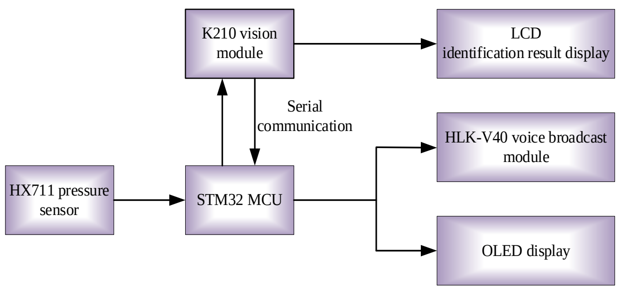


Fig. 1. The system hardware structure of vegetable identification

A. Deep Separable Convolution

The vision processing module adopts the lightweight neural network model MobileNet built by DSC [10]–[15]. This model has carried out a lot of experiments in terms of resource and accuracy tradeoffs. Compared with other popular ImageNet classification models, it has more powerful performance and occupies less memory space, which meets our design requirements.

DSC includes two-step convolution, DC (discrete convolution) and PC (pointwise convolution). The first step is to convolute each input channel with a filter, then input the results calculated by DC to PC for convolution kernel operation, and finally obtain the results. This operation method can effectively reduce the calculation amount and model parameters, so as to obtain more abstract features [9]. MobileNet uses Batchnorm and ReLU nonlinear activation functions for both layers. The depth convolution of each input channel (input depth) can be written as:

$$\hat{G}_{k,l,n} = \sum_{i,j} \hat{K}_{i,j,l} \cdot F_{k+i-1,l+j-1,m}$$

Where $\hat{K}$ is the depth convolution kernel size of $D_K \times D_K \times M$. The m_{th} filter in $\hat{K}$ is applied to the m_{th} channel in F to generate the m_{th} channel of the filtered output feature map $\hat{G}$.

The calculation cost of deep convolution is:

$$D_K \cdot D_K \cdot M \cdot D_F \cdot D_F$$

DC is very effective relative compared to standard convolution, which only filters input channels. In order to generate new features, it needs an additional layer to calculate the linear combination of the depth convolution output through convolution. The cost of DSC is as follows:

$$D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F$$

This is the sum of depth convolution and 1×1 point convolution. By expressing convolution as filtering and combining to reduce the calculation amount :

$$\frac{D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F}{D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F} = \frac{1}{N} + \frac{1}{D_K^{\ 2}}$$

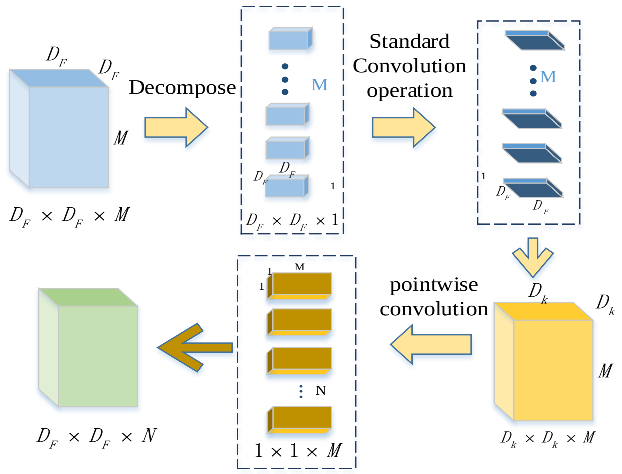


Fig. 2. MobileNet convolution operation process.

B. System Software Design

First, the whole system is powered on. The pressure sensor detects whether vegetables are weighed in real time. If there is a vegetable with a pressure sensor, the MCU drives the vision processing module to collect the vegetable image, and then uses the neural network model to identify it. The vision processing module sends the recognition result to the MCU through the serial port protocol. MCU completes the total price calculation according to the weight and classification results. Finally, the OLED display shows the weight, unit price and total price. The voice broadcasting module broadcasts the name, unit price and total price of vegetables. The display screen displays vegetable images and recognition results.

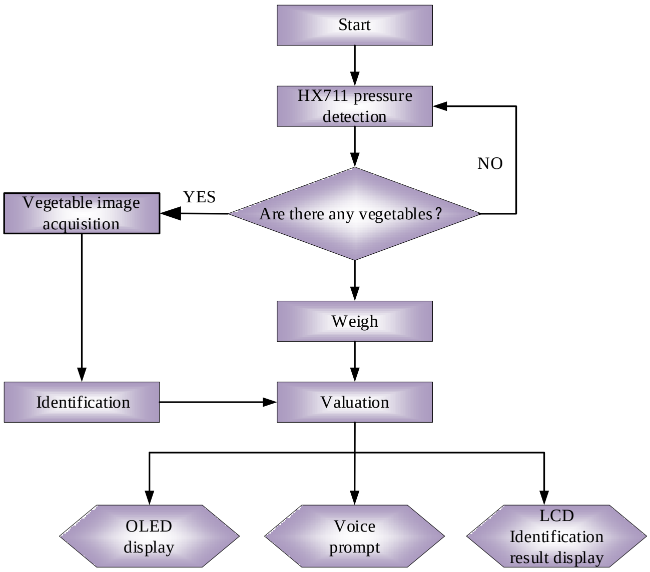


Fig. 3. The software flow of vegetable identification

C. *System Hardware Design*

Fig. 4 is the physical hardware, which contains main control, visual processing, voice announcements and OLED display 4 parts.

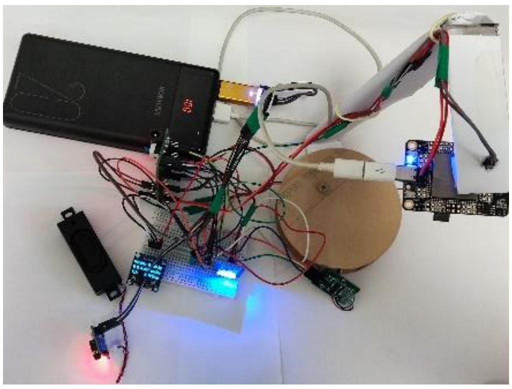


Fig. 4. The physical hardware diagram

1) *Main Control*

We use STM32 micro control board as the main control core. It is a 32-bit microcontroller based on Cortex-M3 core. The hardware adopts the smaller LQFP48 package, which has better performance than the 8-bit MCU, which meet the design requirements of the system.

2) *Visual Processing*

This module uses the open source AI development kit M1 Dock, which integrates Micropython and uses professional chip k210 as the core processing unit. In terms of AI processing, k210 has machine vision capabilities and can perform convolution, batch normalization, activation, pooling and other operations. It has a convolution artificial neural network hardware accelerator KPU, which can perform convolution artificial neural network operations with high performance, and has better low-power visual processing speed and accuracy.

3) *Voice Announcements*

In this module, we use TTS text as speech broadcasting chip, and the chip model is HLK-V40. In addition to voice to text broadcasting, HLK-V40 can realize voice broadcasting of data from serial port and voice broadcasting of data from network. Because HLK-V40 integrates a stereo CODEC with low power consumption and low noise and a high-efficiency stereo amplifier PA in the module, our system not only realizes high-quality sound quality, but also provides programmable control function. This greatly reduces the number of system components and material costs, and ensures high fidelity audio output.

4) *OLED Display*

We used a 0.96 inch display screen, that is, an organic light emitting diode. It has many advantages such as high contrast, wide viewing angle and fast reaction speed. Its resolution is 128×64. Each pixel is an LED with a wide viewing angle greater than 160° . The normal display power consumption is only 0.06W

III. Experimental Environment and Results Analysis

A. *Experimental Environment and Setting*

In order to test our system, five kinds of vegetables, including Cucumber, Tomato, Potato, Cabbage and Radish, were identified. We use MaixPy to debug and design the visual processing module program. The upper left of the interface is a program writing window, and you can debug the program through the serial terminal at the lower part. On the upper right of the interface is the frame buffer of the image captured by the camera, which displays the same content as the LCD. The lower right is the RGB color space of the image, and the value of each channel is $0 \sim 255$. The input image size is 224×224.

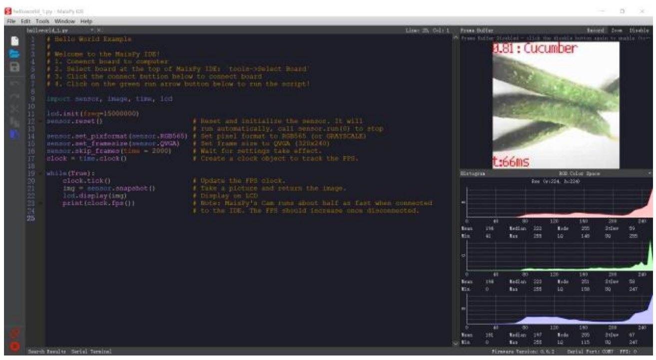
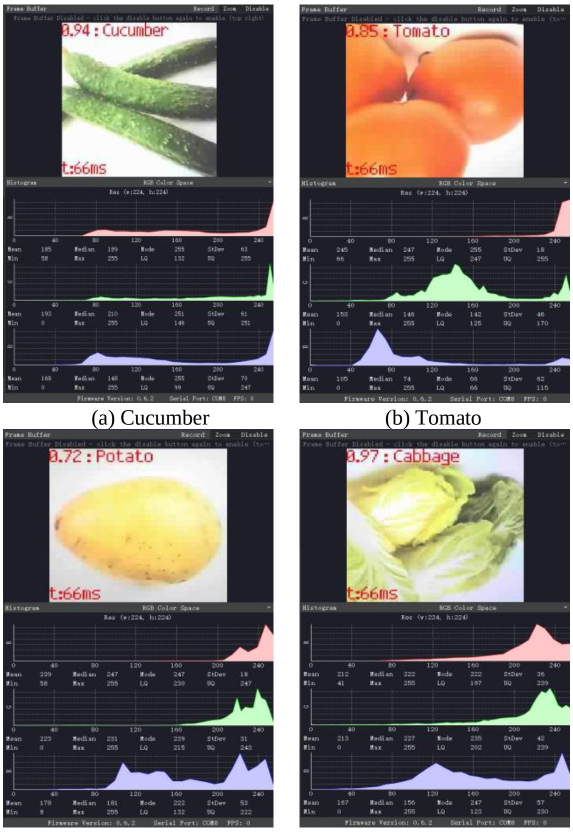


Fig. 5. Maixpy software interface

Fig. 6 is an example of recognition of cucumber, tomato, potato and cabbage respectively. That is, for (a), there is a 94% probability to identify as cucumber, for (b), there is a 85% probability to identify as tomato, for (c), there is a 72% probability to identify as potato, and for (d), there is a 97% probability to identify as cabbage. In addition, the recognition time of these four vegetables is 66ms.



(a) Cucumber (b) Tomato (c) Potato (d) Cabbage

Fig. 6. Identification of four vegetables.

B. Experimental results

In practical application scenarios, due to occlusion and other environmental factors, the brightness of the image captured by the camera will be affected. Therefore, we conducted a recognition test on vegetables under different brightness environments. The brightness of each kind of vegetables was set to three types: Dark, Normal and Bright, and each group had 20 tests. We used the accuracy as the model evaluation index. Accuracy can be derived from the confusion matrix.

TABLE I. CONFUSION MATRIX

	Positive	Negative
Predict as positive	TP(True Positive)	FP(False Positive)
Predict as negative	FN(False Negative)	TN(True Negative)

Accuracy is calculated as follow:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

TABLE II. THE VEGETABLES RECOGNITION RESULTS UNDER DIFFERENT BRIGHTNESS

Vegetables	Brightness	Time (ms)	Accuracy(%)
Cucumber	Dark	66.05	85
	Normal	65.85	100
	Bright	65.80	100
Tomato	Dark	65.90	100
	Normal	66.10	100
	Bright	66.05	100
Potato	Dark	66.00	100
	Normal	66.10	100
	Bright	65.95	100
Cabbage	Dark	66.10	100
	Normal	65.95	100
	Bright	66.05	100
Ternip	Dark	65.90	80
	Normal	65.90	95
	Bright	66.00	100
Total			97.33

Table II shows that five kinds of vegetable identification under different brightness conditions, the recognition probability of each vegetable gradually increased with the increase of brightness. When the brightness is bright, all vegetables can be recognized successfully. When the brightness was dark, only cucumber and white radish could not be recognized correctly, with accuracy of 85% and 80% respectively. In all 300 groups of tests, the overall accuracy can reach 97.33%, and the running time is about 66ms.

From Table II, it can be seen that this system can achieve the classification task well, and the accuracy rate is about 97.33%, which meets the expected design requirements. In practice, the system can guarantee normal operation in most cases because of the fewer parameters and computations of MobileNet. However, due to the reasons for light, image acquisition bias results in recognition failure. How to reduce the impact of light is the next research direction.

IV. CONCLUSIONS

This paper designs am intelligent vegetable recognition system based on MobileNet, including software system design and hardware system design. The hardware system includes the main control core, visual processing module, pressure sensor, voice broadcasting module and display module. The software system mainly includes image acquisition module, processing and analysis module, serial communication module, training module and recognition module. The experimental results show that the accuracy of five kinds of vegetables can reach 97.33% under three kinds of light conditions: dark, normal and bright. By completing the above work, the research of vegetable species recognition algorithm based on deep learning theory can lay the foundation for intelligent vegetable species recognition, which has important practical significance.

REFERENCES

[1] Arivazhagan, S. , Shebiah, N. , Nidhyanandhan, S. , Ganesan, L. "Fruit recognition using color and texture features," journal of emerging trends in computing & information sciences. 2010.

[2] Rocha, A. Hauagge, D. C. Wainer, J. Goldenstein, S. "Automatic fruit and vegetable classification from images," in *Comput Electron Agric*, 2010, vol. 70,no.1, pp. 96-104.

[3] I Dubey, S. R. Jalal, A. S. "Robust Approach for Fruit and Vegetable Classification," procedia engineering, 2012.

[4] Dubey, S. R, A. S. Jalal . "Detection and Classification of Apple Fruit Diseases Using Complete Local Binary Patterns." In *ICCCT*-2012, 2012.

[5] S. C, N. Manasa, V. Sharma, N. K. A. A., "Vegetable Classification Using You Only Look Once Algorithm," in *ICon-CuTE*, 2019, pp. 101-107, doi: 10.1109/ICon-CuTE47290.2019.8991457.

[6] Ghosh T , Abedin M , Reza S , et al. "Bangla Handwritten Character Recognition using MobileNet v1 Architecture," Bulletin of Electrical Engineering and Informatics, 2020, 9(6).

[7] Lei, Y. Baowei, D. U. Qian, J. Feng, Z. "Research on Ear Recognition Based on SSD_MobileNet_v1 Network." in *CAC*, 2020.

[8] Prasetyo, E. Purbaningtyas, R., Adityo, R. D., Suciati, N., Fatichah, C. "Combining MobileNetV1 and Depthwise Separable convolution

bottleneck with Expansion for classifying the freshness of fish eyes," in *IPA*, 2022, vol. 9,no. 4, pp. 485-496.

[9] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H. "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," 2017.

[10] Venkatesh, N. Y, S. U. Hegde, S. S, "Fine-tuned MobileNet Classifier for Classification of Strawberry and Cherry Fruit Types," in *ICCCI*, 2021, pp. 1-8, doi: 10.1109/ICCCI50826.2021.9402444.

[11] P. S. P. Kavyashree, M. El-Sharkawy, "Compressed MobileNet V3:A Light Weight Variant for Resource-Constrained Platforms," in *CCWC*, 2021, pp. 0104-0107.

[12] Y. Sun, J. Zhang, C. Han, "A flower recognition system based on MobileNet for smart agriculture," in *ICFTIC*, 2021, pp. 712-717.

[13] J. C. V. Gómez, A. P. Z. Incalla, J. C. C. Perca, D. I. M. Padilla, "Diferentes configuraciones para MobileNet en la detección de tumores cerebrales: Different configurations for MobileNet in the detection of brain tumors," in *ICALTER*, 2021, pp. 1-4.

[14] J. Liu, "VGG, MobileNet and AlexNet on Recognizing Skin Cancer Symptoms," in *IWECAI*, 2022, pp. 525-528.

[15] S. Bouguezzi, H. Faiedh, C. Souani, "Slim MobileNet: An Enhanced Deep Convolutional Neural Network," in *SSD*, 2021, pp. 12-16.