A Real-time Image Recognition System Based on Improved Jacintonet Convolutional Neural Network

Shixing Chen¹, Hongfang Yuan¹, Xi Cao¹ and Xiang Li¹
¹ College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China
²E-mail: caoxi@mail.buct.edu.cn

Abstract. With the emergence and development of new automation industries such as unattended supermarkets and smart picking orchards, the demand for real-time systems based on embedded platforms is increasing day by day. Heterogeneous multi-core processors are widely used in modern integrated circuit design due to their advantages of low power consumption and high parallelism. More and more real-time systems are implemented on heterogeneous multi-core platforms. Based on the heterogeneous multi-core embedded system, the network parameters and architecture of jacintonet model are improved, and a real-time system for fruit image recognition is realized by using the improved network. A small sample data set is used to train the modified network and the trained network is imported into the system for testing. The result shows that the improved jacintonet network can run well on heterogeneous multi-core system and has the same recognition performance as the original network.

1. Introduction
In recent years, with the wide application of electronic payment technology and the maturity of image recognition technology, more and more self-service industry has appeared in people's vision. With the continuous development of deep learning, convolutional neural networks have shown excellent performance in image fields such as target detection and face recognition [1]. Compared with other traditional image processing methods, convolutional neural networks have the following advantages [2]: (1) better image feature extraction ability. Convolutional neural networks have the characteristics of local perception. Each neuron is not need to perceive all pixels in the image. It only perceives local pixels in the image, and then combines these local information at a higher level to obtain the full representation of the image information; (2) End-to-end image processing capabilities. The convolutional neural network can automatically extract the features of the image, and with the deepening of the network layers, the recognition accuracy can be further improved; (3) Weight sharing. One image uses the same parameters of the convolution kernel, which not only reduces the number of weights, but also greatly reduces the complexity of the model.

As the key to many systems, fruit image recognition technology has been widely used in various fields. In the field of smart agriculture, through the recognition of fruit images, precise cultivation and automatic picking [3] can be achieved; in the field of digital medical treatment, fruit image recognition technology can be used to analyze the nutritional composition of fruits, so as to help patients to make a reasonable diet. In addition, fruit image recognition technology is also applied to automatic counting and automatic weighing [4] in supermarkets, bringing convenience to people's daily lives.

Heterogeneous multicores processors are usually designed for certain applications, and different cores
handle programs or tasks that they are good at. Therefore, heterogeneous multicore processors have more flexible and efficient processing mechanism [5], and can make better use of chip area to reduce power consumption. Based on the embedded platform of AM5708 heterogeneous multi-core processor, this paper implements a real-time system of fruit image recognition with the improved jacintonet network. AM5708 is a multi-core heterogeneous processor composed of ARM, DSP and other units. In the process of image recognition, ARM is responsible for task allocation and system control, while DSP completes data processing and calculation. The digital signal processor (DSP) contained in this heterogeneous processor not only has a strong processing capability for fixed-point data, but also enhances the operational capability of floating point, matrix and fixed-point vector, which ensures the real-time recognition of the system. In addition, the heterogeneous embedded system supports TIDL, a deep learning framework optimized for the convolutional neural network, which further improves the performance of the improved jacintonet network on the system.

The contribution of this paper are as follows;
• An improved jacintonet network is proposed. In order to make better use of the computing performance of heterogeneous multi-core embedded platforms, the size of some convolutional layer’s weight is modified, and the stride of the weights of convolutional layers is changed. The modified jacintonet network has a good sparse representation ability, which can well learn the characteristics of sample categories of small data sets.
• A real-time system of image recognition is designed based on heterogeneous embedded platform. Based on the AM5708 processor, a real-time system for fruit image recognition is designed and implemented on the multi-core heterogeneous embedded system with the improved network. After the image information collected by the camera is identified by the system, the recognition result is output.

The paper is organized as follows. Section 2 gives the design process of the model, including the introduction of the original jacintonet network, the improvement process of jacintonet network parameters and architecture, and the production of data set. Section 3 provides the experimental results and demonstration analysis of the system test. The comparison of the recognition performance between original and improved network is also given in this section. Section 4 concludes the paper and simply analyzes the application prospects of the system.

2. Related work
Convolutional neural network (CNN) is a typical multi-layer neural network, which takes the image as the input, and outputs the recognition result after the input image passes through multiple convolutional layers, pooling layers and full connection layers. A typical convolutional neural network structure [6] is shown in figure 1.

![Typical structure of CNNs.

2.1 Jacintonet network
The research shows that the increase of convolutional layers can effectively improve the network's learning ability and the accuracy of image recognition, but it will also bring the network degradation and the disappearance of the gradient. The residual network (ResNet) replaces the convolutional layer in the convolutional neural network with the residual unit [7], which solves the problem of network degradation and gradient disappearance as the number of layers deepens. The jacintonet network design
idea is derived from ResNet 10. Compared with the ResNet 10 network, jacintonet has made the following changes: (1) Removed the identity connection structure because the shallow network does not make sense to use the identity connection structure [8]. (2) Added groups of 4 to every alternate layer to reduce network complexity and data bandwidth [9]. (3) Max pooling is used to instead of strides.

2.2 Improved jacintonet network

The platform chosen for this test is embedded system based on AM5708, and the system framework is shown in figure 2. AM5708 is a multi-core and heterogeneous processor, in which the microprocessor (MPU) is mainly responsible for image reading and result output, the digital signal processor (DSP) is responsible for image information processing and calculation, and the image processing unit (IPU) is responsible for assisting the microprocessor to control the system. The system can collect 3-channel color images and supports TIDL, a deep learning framework optimized for convolutional neural network. The heterogeneous architecture of the core processor also provides the possibility for fast processing of image data.

The original and improved jacintonet network parameters are given in Table 1. The input to the improved network is 3 channels of the image of resolution 224*224. The 512-dimensional feature map with the size of 7*7 is entered into the full connection layer after passing through the maximum pool layer. Finally, the recognition probability of 10 kinds of fruits is obtained through the softmax function. Compared with the original network, the improved jacintonet network parameters and architecture are modified as follows: (1) The size of the convolution kernel of all convolutional layers is adjusted to 3*3, and the stride is set to 1, which ensures that the size of the image remains the same after passing through the convolutional layer. (2) All pooling layers use maximum pooling. Compared with average pooling, maximum pooling can reduce the mean deviation caused by the parameter error of the convolutional layer and retain more texture information of the image. In addition, the use of the maximum pooling layer reduces the image features and network parameters, which further reduces the complexity of the network. (3) Groups of 4 are used in every alternate layer to reduce the complexity and data bandwidth. The improved jacintonet network has strong sparse representation ability [10] and can well learn the characteristics of sample categories of small data sets.

![Figure 2. The framework of system.](image-url)

| Layer No | Layer type | Kernel size original | Stride original | Kernel size improved | Stride improved |
|----------|------------|----------------------|----------------|---------------------|----------------|
| 1        | Conv, Relu | 5                    | 2              | 3                   | 1              |
| 2        | Conv, Relu | 3                    | /              | 2                   | 1              |
| 3        | Maxpool    | 2                    | 2              | 2                   | 2              |
| 4        | Conv, Relu | 3                    | /              | 1                   |                |
| 5        | Conv, Relu | 3                    | /              | 1                   |                |
| 6        | Maxpool    | 2                    | 2              | 2                   |                |
| 7        | Conv, Relu | 3                    | /              | 1                   |                |
2.3 Data set
In this paper, the common fruits in unattended supermarkets are used to make training samples, a total of 10 types of fruits are selected to make a dataset of pear, apple, banana, cherry, peach, lemon, orange, avocado, tomato and kiwi. Ten fruits are chosen in different sizes and shapes for each type, and each fruit is 10cm, 20cm and 30cm away from the camera. Besides, all pictures are taken at 10 different angles at the same distance, thereby obtaining 3000 original sample pictures. 3,000 original pictures are translated, scaled, flipped, and Gaussian blurred, and finally 12,000 pictures are obtained as training samples for the network. Figure 3 shows an example of some pictures in the data set, taking a lemon as an example.

![Figure 3. Examples of data set.](image)

3. Results and discussions
The original network and the improved network are trained with the same data set in this test, and the trained network is imported into the heterogeneous multi-core embedded system for testing. The environmental conditions during the test were the same as those of the samples. This section presents the test results and analysis of the system.

3.1 Results
In the actual test process, ten fruits are chosen in different sizes and shapes for each type in the data set, and 5 different angles were randomly selected at a distance of 10 cm, 20 cm, and 30 cm from the camera for testing. A total of 150 tests are performed on each fruit. Figure 4 shows the identification process of different types of fruits. The test results and accuracy are given in figure 5.

As can be seen from the comparison of test accuracy in figure 5, the original and improved network recognition performance of fruits with obvious characteristics is almost the same, while for fruits with highly similar size, shape and color, such as pear and apple, cherry and cherry, the network recognition accuracy with improved network is higher than original one. On the one hand, the improved network has stronger sparse representation ability and learning ability for small sample data; on the other hand, the use of maximum pool layer enables the network to retain more texture information in the process of learning, so as to better identify different kinds of fruits with highly similar appearance.
Figure 4. Identification results of banana and pear under different conditions.

Figure 5. Accuracy of original and improved network.

3.2 Discussions
For the convolutional neural network, in the process of image information processing, the convolutional layer will generate a large amount of calculation data, but also consume the Multiply and Accumulate (MAC) of hardware. The modified jacintonet network adjust the size of the weight of the first convolutional layer. Parameters and MAC operation of original and improved network are shown in table 2. As can be seen from the data in table 2, the improved network has significantly reduced the computational data and resource consumption in the convolutional layer, which reduces the complexity of the model. Meanwhile, the reduction of data and MAC saves the storage and DSP resources of the system, which further reduces the cost and power consumption of the system.

| Fruit type  | Original Parameters | MAC | Improved Parameters | MAC |
|------------|---------------------|-----|---------------------|-----|
| pear       | 224*224*3           | $5*5*3*32=2400$ | 224*224*3           | $3*3*3*32=864$ |
| apple      |                     |     |                     |     |
| banana     |                     |     |                     |     |
| cherry     |                     |     |                     |     |
| peach      |                     |     |                     |     |
| lemon      |                     |     |                     |     |
| orange     |                     |     |                     |     |
| avocado    |                     |     |                     |     |
| tomato     |                     |     |                     |     |
| kiwi       |                     |     |                     |     |

Table 2. Parameters and MAC operation of original and improved network.
Although the improved network further reduces the complexity of the network compared with the original network, and has better learning ability of small data samples, it still has shortcomings, mainly reflected in the inability to accurately identify different kinds of fruits of similar size, shape and color, such as cherry and cherry, pear and apple. The training set in this paper and the test environment have the same lighting conditions and background. In practical application, the image acquired by the camera will be interfered by lighting, shielding and background color, which will have a certain influence on the recognition result.

4. Conclusion
In this paper, based on the improved jacintonet network, a real-time fruit image recognition system is designed and implemented by using a heterogeneous multi-core processor with AM5708 as the core. The experimental results show that the improved network has better learning ability of small sample data, and the network complexity is further reduced. Due to its advantages of low power consumption, low cost and high parallelism, heterogeneous multi-core processors can be well combined with convolutional neural network, laying a foundation for the development of many new retail industries based on image recognition and target detection.

References
[1] Kong J and Wang X 2018 Study on image recognition based on PCNN and LVQ IEEE International Conference on Automation, Electronics and Electrical Engineering 177- 80
[2] Huang Z T and Zhang Q 2019 Skew correction of handwritten chinese character based on ResNet International Conference on High Performance Big Data and Intelligent Systems 223-7
[3] Jana S, Basak S and Parekh R 2017 Automatic fruit recognition from natural images using color and textures features Devices for Integrated Circuit 620-624
[4] Kuang H, Chan L L H and Yan H 2015 Multi-class fruit detection based on multiple color channels Proceedings of the 2015 International Conference on Wavelet Analysis and Pattern Recognition 1-7
[5] Rafiev A, Al-Hayanni M, Xia F, Shafik R, Romanovsky A and Yakovlev A 2018 Speedup and power scaling models for heterogeneous many-core systems IEEE Transactions on Multi-scale computing systems 4(3) 436-49
[6] Ding R, Su D G, Bai G Q, Xu W, Su N and Wu X J 2019 A fpga-based accelerator of convolutional neural network for face feature extraction IEEE International Conference on Electron Devices and Solid-State Circuits
[7] Lu Z, Jiang X D and Kot A 2018 Deep coupled ResNet for low-resolution face recognition IEEE Signal Processing letters 25(4) 526-30
[8] He K M, Zhang X Y, Ren S Q and Sun J 2016 Deep residual learning for image recognition IEEE Conference on Computer Vision and Pattern Recognition 1-12
[9] Mathew M, Swami K D P K and Nagori 2017 Sparse, quantized, full frame CNN for low power embedded devices IEEE Conference on Computer Vision and Pattern Recognition Workshops 328-36
[10] Viswanath P, Nagori S, Mody M, Mathew M and Swami P 2018 End to end learning based self-driving using jacintoNet IEEE 8th International Conference on Consumer Electronics