

Design of automatic sorting system based on capsule neural network

Yanlin Ju

School of Information Science and Technology, East China University of Science and Technology, Shanghai, China

22013199@mail.ecust.edu.cn

Abstract. This paper proposes a vegetable disease classification system based on Capsule Neural Network (CapsNet), which uses the visual recognition mechanism of deep learning algorithms to improve the efficiency and accuracy of agricultural disease recognition. The CapsNet algorithm outperforms traditional Convolutional Neural Network (CNN) in few-shot learning and generalization due to its unique capsule structure and dynamic routing mechanism. Experimental results show that the proposed system performs well in the accuracy, robustness, and model interpretation of image recognition, and can accurately identify diseases even in the case of poor image quality or noise and occlusion. In addition, the embedded system of this experiment includes an image acquisition module, an algorithm core module, and a logic control module, which realizes the automation of the whole process from image acquisition to recognition result display. This study also discusses the scalability and adaptability of the system, as well as the future application prospects in agricultural intelligence and automation. It is expected that CapsNet will play a more critical role in artificial intelligence, especially in agrarian automation and smart agriculture.

Keywords: capsule neural network, embedded system, visual recognition, smart agriculture.

1. Introduction

1.1. Background and Significance

As the cornerstone of human civilization, agriculture not only ensures food security but also profoundly impacts economic development, environmental protection, and cultural heritage. With the progress of science and technology, especially the rise of agricultural mechanization and artificial intelligence, agricultural modernization is gradually being realized.[1]

In terms of agricultural mechanization, although large machinery improves production efficiency, in countries with complex terrain and scattered land such as China, small machinery has become a new trend due to its flexibility and adaptability, which can better meet diversified agricultural needs. With the rapid development of artificial intelligence technology, especially the breakthrough of deep learning in the field of image recognition, a new solution has been provided for agricultural modernization. By integrating deep learning technology into small agricultural machinery, it is possible to realize the automatic identification, harvesting, and packaging of crops such as vegetables, which not only greatly

improves the efficiency and accuracy of operations, but also reduces labor costs and improves the quality and market competitiveness of agricultural products.

This visual recognition system based on deep learning can quickly and accurately identify the maturity, rot degree, and other characteristics of vegetables, so as to realize fine management and automatic operations. In this paper, the capsule neural network is applied to the embedded system to develop a machine for field recognition operation, which helps reduce costs and improve efficiency and product quality. The system can be adjusted and optimized according to different crops and environments, and promote the development of agriculture in the direction of more intelligent and automated.

1.2. Related Work

In this paper, the capsule neural network is combined with the embedded system to improve the efficiency of agricultural harvest. The aim is to capture the actual harvest and crop images for recognition through the capsule neural network and make full use of the advantages of small size, low power consumption, and high flexibility of the embedded system. [2]Both of them have been fully explored in the front-end research at home and abroad.

1.2.1. Research and development of capsule neural network

Capsule Network (CapsNet) is a deep learning architecture proposed by Geoffrey Hinton and his team, which aims to overcome the insufficient processing of spatial relationships in image recognition by traditional convolutional neural networks (CNNs). CapsNet improves the generalization ability and flexibility of the model by retaining the location information of features and introducing a dynamic routing mechanism. [3] Compared with traditional CNNs, CapsNet is able to better understand the relationships between data, especially when dealing with object recognition tasks with different perspectives and deformations. In Figure 1, the right picture is not a face but has the elements needed for a face, so it is possible that the CNN activates the judgment that it is a face by having the object, thus making the result wrong.

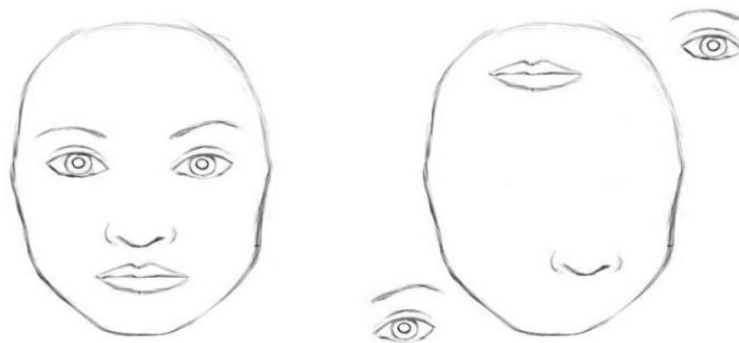


Figure 1. Normal faces and abnormal faces

Since the concept was proposed in 2010, the research on CapsNet has developed rapidly, including matrix capsule network, multi-scale CapsNet, HitNet, multi-channel CapsNet, and complex-valued CapsNet. They optimize the performance and application range of CapsNet using different methods. [7]With the further development of the technology, CapsNet is expected to play a more critical role in the field of artificial intelligence.

1.2.2. Research and Development of Embedded System

For the early foreign countries, the embedded system has experienced the development from single chip microcomputer to microcontroller since the microcomputer era, and with the emergence of SoC technology and the adoption of real-time operating systems, its integration and performance have been

significantly improved. In the 21st century, with the improvement of microprocessor performance and power efficiency, embedded systems have been widely used in the field of the Internet of Things. Future trends, including the integration of software and hardware, edge computing, and intelligent development, will further promote the application of embedded systems in many industries.

China has made significant progress in the field of embedded systems, and the market size continues to expand[8]. Domestic embedded operating systems, such as HUAWEI Lite OS, SylixOS, AliOS Things, and RT-Thread, have been applied to a large scale in several fields, which has promoted the development of domestic software. With a market size of nearly 100 billion yuan in 2019, it has become an important part of the industrial software market. The Chinese government's support and national strategies such as 2025 for developing embedded systems in China provide a good policy environment. With the improvement of new applications and industrialization requirements, embedded system software is developing towards high credibility, adaptability, and component-based, and the development environment tends to be integrated, automated, and humanized. [4]The competition in China's embedded software market is fierce, and domestic and foreign enterprises jointly promote technological progress and market development. The embedded system industry in China is expected to continue its growth momentum.

2. The principle of the Capsule Neural Network

The prominent output feature of the capsule neural network is vector input vector output, which better simulates the process of the human brain in 3D modeling, analysis, and identification of the same object from different angles. Compared with the traditional CNN to expand the training data set to achieve similar functions, the capsule neural network uses the dynamic routing algorithm through the establishment of the capsule layer and important parameters for effective simulation.

2.1. Network model structure and operation

2.1.1. Model Structure

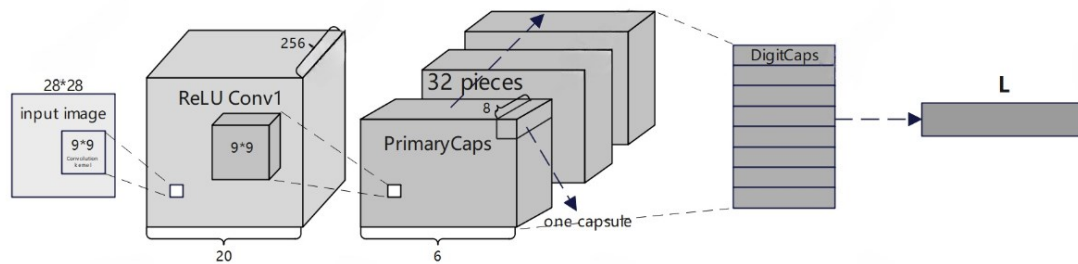


Figure 2. Schematic diagram of a single-capsule neural network

Convolutional layer: The job of this layer is to detect essential features in a 2D image. Unlike traditional CNNs, the convolutional layers of CapsNet output vectors instead of scalars. These vectors contain not only the feature's existence probability but also the feature's spatial pose information. Such vector output provides the basis for dynamic routing. In CapsNet, the vector extracted by the convolutional layer can be represented as \vec{u} .

PrimaryCaps layer: This layer is located after the convolutional layer and its job is to take the output of the features detected by the convolutional layer and generate the combination of the input features, further process the features extracted by the convolutional layer, and create the capsule. [9]Each capsule unit in the PrimaryCaps layer is connected to multiple capsules in the DigitCaps layer, and the weight matrix determines the initial strength of these connections. The coupling coefficients are updated during the dynamic routing process to reflect the correlation between PrimaryCaps and DigitCaps.

DigitCaps layer: This layer consists of a set of capsules, where each capsule represents a category.[10] The output vector of each PrimaryCaps is connected with multiple capsules in the DigitCaps layer and goes through an iterative dynamic routing process to activate the most relevant DigitCaps. At the same time, the activation of each capsule indicates whether a certain feature exists the direction of the feature, and the probability of its existence.[14]

2.1.2. Calculation process

For a vector, the complete CapsNet algorithm goes through the following process

First, the weights are assigned, which can be expressed as the following formula:

$$\hat{u}_{ji} = W_{ij} * u_i$$

\hat{u}_{ji} is the output vector of the PrimaryCaps layer; u_i is the output vector of the convolutional layer; W_{ij} is the weight.

Secondly, the coupling and summation can be expressed as the following formula:

$$s_j = \sum_i c_{ij} * \hat{u}_{ji}$$

s_j is the output vector of the DigitCaps layer; c_{ij} is the coupling coefficient whose numerical magnitude is related to the pose between layers relative to each other; \hat{u}_{ji} is the output vector of the PrimaryCaps layer.

For the activation function, it can be expressed by the following formula:

$$v_j = \frac{\|s_j\|^2 \cdot s_j}{(1 + \|s_j\|^2) \cdot \|s_j\|}$$

v_j is the output vector. The function of the main body of the formula is similar to that of a traditional CNN, but the output is still a vector.

Therefore, Figure 3 is an overview of how capsule neural networks work.

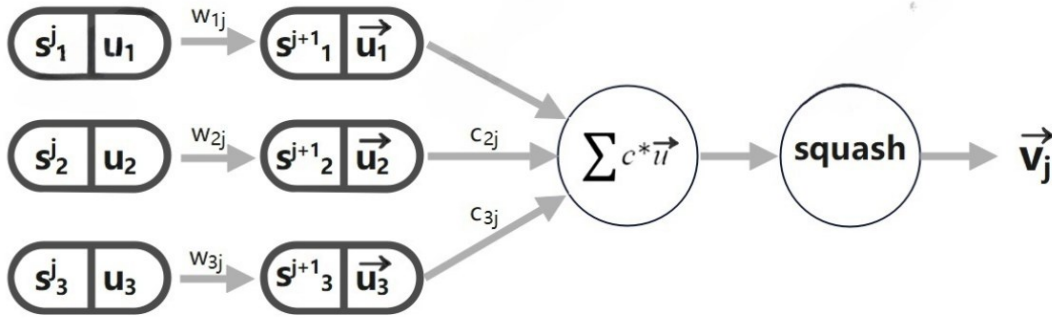


Figure 3. An overview of how capsule neural networks work

2.2. Dynamic Routing Algorithm

The capsule network takes a similar input as a fully connected neural network, namely a linear weighted summation, but it adds an additional coupling coefficient C to the linear summation phase. [11] C is calculated according to the following formula:

$$c_{ij} = \frac{e^{b_{ij}}}{\sum_k e^{b_{ik}}}$$

Therefore, for the solution of C , b is needed, by the following formula:

$$b_{ij} \leftarrow b_{ij} + \hat{u}_{j|i} \cdot v_j$$

The initial value of b is 0. Therefore, in the process of forward propagation to calculate S , W is designed to be a random value, b is initialized to zero and can be obtained [12], \hat{u} is the output of the capsule network of the previous layer. With these values, the S of the next layer can be obtained further. Therefore, the idea of the dynamic routing algorithm is simply to update b to update c .

Vector point multiplication's result is positive or negative, depending on the direction of the two vectors. For the point formation operation of $\hat{u}_{j|i}$ and v_j , when the result is positive, it means that the direction of the two vectors is similar, and the update result of b will become larger, which indicates that the coupling coefficient is high, which means that the two vectors match well. On the contrary, if the dot product result is negative, the update result of b will become smaller and the coupling coefficient is low, indicating that the two vectors match poorly. [13] Through the iterative process, a parameter C can be determined.

To sum up, the whole dynamic routing algorithm can be summarized as, first, initialize all b to zero, and then start the iteration. Each iteration by Softmax to find the C value first, and combine W , C , and \hat{u} to get S by linear summation. Then enter S to activate the function (Squashing) to get V , and finally use \hat{u} and v_j to complete the update of the b value. After all the calculations are finished, the next iteration is started, and the number of iterations is set to 3.[5]

2.3. Loss function establishment

$$L_c = T_c \max(0, m^+ - \|v_c\|)^2 + \lambda(1 - T_c) * \max(0, \|v_c\| - m^-)^2$$

The meanings are as follows:

$T_c \max(0, m^+ - \|v_c\|)^2$: The loss term characterized by the correct number capsule. m^+ is a constant; $\|v_c\|$ is the L2 norm of the capsule vector v_c . When $\|v_c\|$ is less than m^+ , the loss is zero, otherwise the loss is $T_c \max(0, m^+ - \|v_c\|)^2$, which means the fitting is not appropriate at this time.

$\lambda(1 - T_c) * \max(0, \|v_c\| - m^-)^2$: Loss term for the wrong number capsule. λ is a regularization parameter, T_c is a threshold, $\|v_c\|$ is the L2 norm of the incorrect capsule, and m^- is another constant. When m^- is greater than $\|v_c\|$, the loss is zero, otherwise the loss is $\lambda(1 - T_c) * \max(0, \|v_c\| - m^-)^2$.

3. Algorithm establishment

3.1. Dataset

To test the performance of CapsNet, we use the archive(3) dataset. The dataset consists of 2932 images of size 512*512*3, of which 2889 images are training images and 53 images are testing images. All of the images are divided into three classes, corresponding to three kinds of white leaf disease, of which 1794 pieces are for Backmoth, 333 are for Leafminer, and 752 are for Mildew.

3.2. Comparison of algorithms

This article compared the traditional CNN and CapsNet and tested CapsNet algorithm on white leaf lesions. Firstly, the traditional cnn algorithm adopts the Input -> [Flatten] -> fc1 (Linear Layer) -> fc2 (Linear Layer) -> Output structure.

The training accuracy and validation accuracy of the traditional cnn algorithm are as follows:

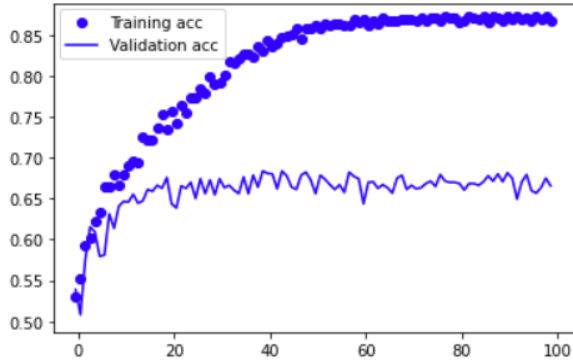


Figure 4. Training and validation accuracy

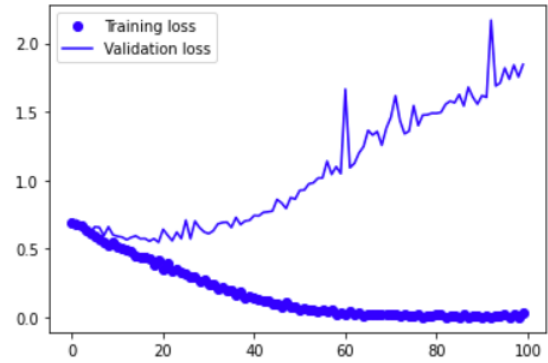


Figure 5. Training and validation loss

It can be seen from the above pictures that the traditional CNN algorithm has a poor fit between prediction and theory, and the accuracy does not reach more than 0.9.

For the CapsNet algorithm, convolutional layer: The convolutional layer has 256 kernels of size $9 \times 9 \times 1$ with a step size of 1. PrimaryCaps layer 32 "Lord capsule", each capsule shape is the output of $6 \times 6 \times 8$ tensor. The final output of this layer is $6 \times 6 \times 8 \times 32$. DigitCaps layer, each capsule receives a $6 \times 6 \times 8 \times 32$ shape vector and outputs a 16×10 matrix.

The training accuracy and validation accuracy are obtained by experiments as shown in Figures:

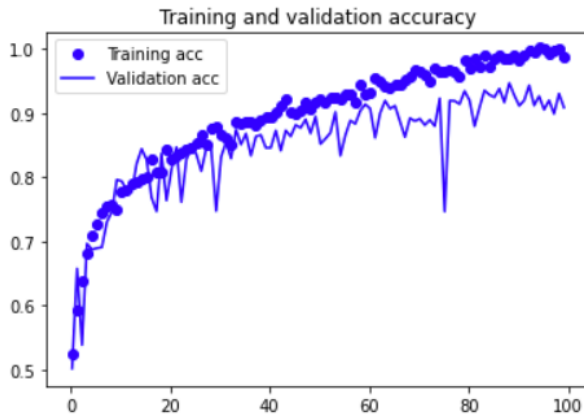


Figure 6. Training and validation accuracy

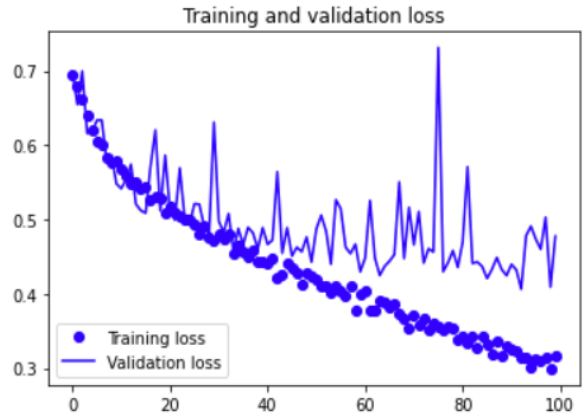


Figure 7. Training and validation loss

By comparing available conclusions:

- Stronger generalization ability: CapsNet algorithm has better generalization, and maintains a higher recognition accuracy.
- For small sample learning adaptability: CapsNet when dealing with small sample data sets, reduces the dependence on large-scale training sets.
- Robustness: in the case of poor image quality, for the noise in the input data and keep out CapsNet has higher robustness. Namely, the image state requirements are low, with less data enhancement dependence.
- Explanatory: CapsNet output provides information about identifying causes, namely, further analysis of the type and degree of pathological changes.

4. Embedded system design

4.1. Style and spacing

Aiming at the combination of CapsNet algorithm and embedded system, this paper selects two embedded motherboards, Raspberry Pi and Arduino, to make the system easy to expand and have powerful processing power. Specifically, the system is divided into a mechanical hardware end, image acquisition module, algorithm core module, and logic master control module. The system mode is shown in Figure 8.

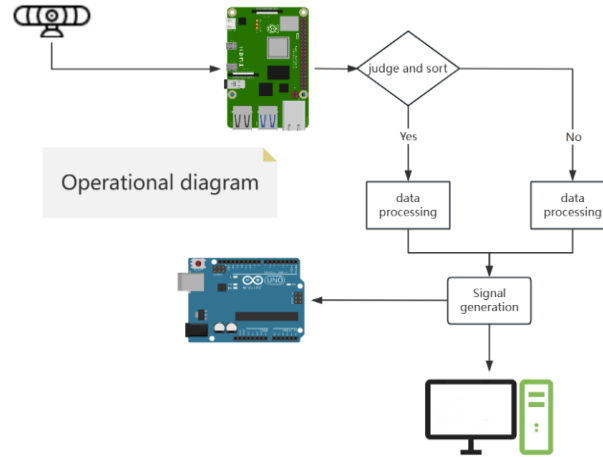


Figure 8. Schematic diagram of embedded system construction

4.1.1. Mechanical Hardware Side

The Raspberry Pi is less scalable, so Arduino is used to connect the mobile parts. [15] This design avoids the disadvantages of two kinds of embedded chips and reduces the cost. The Arduino chip is responsible for moving and sorting the DC motor and servo used in the work.

4.1.2. Image Acquisition Module

We choose the Raspberry Pi Pico for the hardware configuration to undertake image extraction and recognition judgment, and transmit the signal to the subsequent module. The system uses Camera Module 3 and its parameters are as follows:

Table 1. Camera's parameters

Name	Size	Weight	Still resolution	Sensor
Camera Module 3	25 x 24 x 11.5 mm	4g	11.9 Megapixels	Sony IMX708

Table 2. Camera's parameters

Sensor resolution	Sensor image area	Pixel size	Focal length	Maximum exposure times
4608 x 2592 pixels	6.45 x 3.63mm	1.4 μm x 1.4 μm	4.74mm	112s

For software configuration, we for Raspberry PI, configure SSH network service. motion monitoring software was used to obtain images in real time and transmit them to the algorithm core module.

4.1.3. Algorithm core module

The key algorithm using the trained CapsNet algorithm can achieve the following functions.

Image processing: Implement image processing algorithms, such as noise reduction, edge detection, feature recognition, etc.

Machine learning: Integrating machine learning models for object recognition, classification, or anomaly detection.

Data Analytics: Analyze image data to extract useful information and generate reports as needed.

4.1.4. Logic master control module

It integrates the control logic of the system, including task scheduling, status monitoring, exception handling, etc. At the same time, it takes into account the user interaction interface and provides the system status display, parameter setting, and result display functions. In addition, the communication interface between the module and the algorithm core module, and the mechanical hardware end are designed to ensure that the system can exchange data with these modules and control other functions of the machine

4.2. Overview of the identification process

The visual recognition system composed of Raspberry PI combined with a CSI camera captures real-time images by starting the system and initializing the camera, followed by image preprocessing and feature extraction to enhance image quality and highlight key information. The system uses motion detection and object recognition technology to locate and classify objects in the image, and then annotates the image and records the recognition results. These results are displayed through the user interface, and remote notifications or system optimization can be performed as needed. The whole process also includes error handling and logging to ensure the stability and maintainability of the system, making Raspberry PI a powerful and flexible visual recognition platform.

5. Conclusions

5.1. Main results

This study successfully developed an intelligent recognition system for vegetable disease classification by combining a capsule neural network (CapsNet) and an embedded system, improving agricultural disease recognition efficiency, accuracy, and scalability.

The results of this paper are remarkable, mainly reflected in the following aspects: First, the CapsNet algorithm significantly outperforms traditional CNN in performance, showing higher training and validation accuracy and stronger generalization ability, especially when dealing with small sample datasets, reducing the dependence on large-scale training data. Secondly, CapsNet significantly improves the robustness of the network through its dynamic routing algorithm, and can accurately identify diseases even in the case of poor image quality. In addition, CapsNet enhances the interpretability of the model and provides deeper information for lesion analysis. At the same time, CapsNet has low requirements on the state of the input image and good tolerance to image scaling, rotation, and partial occlusion, which reduces the dependence on data augmentation techniques and simplifies the preprocessing process. Finally, this study successfully combined the CapsNet algorithm with the embedded system to construct an automated vegetable disease recognition system, which realized the automation of the whole process from image acquisition to recognition result display and provided an efficient and reliable technical solution for disease management in modern agricultural production.

5.2. Application prospect and future work

The intelligent recognition system for vegetable diseases developed in this study is highly scalable and adaptable and can be adjusted to different crops and environments. It reduces costs, improves efficiency by reducing manual intervention, and helps to ensure product quality and market competitiveness.

Although the current results are encouraging, there are several challenges that need to be addressed in future research. First, the recognition accuracy needs to be improved, and the misjudgment needs to be reduced by expanding and diversifying the training dataset; Second, optimize the performance of embedded systems and explore energy-efficient hardware; Third, expand the application scope to more

crop disease management and growth monitoring; Fourth, integrating sensor data and environmental information, enhancing intelligent decision-making capabilities, and achieving a higher level of automation and precision management. These improvements will provide more solid technical support for agriculture's intelligent and sustainable development.[16]

References

- [1] Hanyu Zhao, Hongyan Chai. Strengthen agricultural mechanization technology innovation Promoting agricultural modernization process [J]. Journal of Jiangsu Agricultural Mechanization, 2012, (4): 14-16. DOI: 10.16271 / j.carol carroll nki JSNJH. 2012.04.011.
- [2] Liu Fo, LI Weisheng. Application of embedded system in Shuguang 4000A cluster Monitoring [J]. Microcomputer Information, 2005, (05): 57-58.
- [3] Kranti K, Joydeep S . A comprehensive survey on emotion recognition based on electroencephalograph (EEG) signals [J]. Multimedia Tools and Applications, 2023, 82 (18): 27269-27304.
- [4] Li Guanghui. Research on Embedded Software Related problems [J]. Information Technology, 2007, (02): 67-68.
- [5] Ran C, Hao S, Zhong-Qiu Z, et al. Global routing between capsules [J]. Pattern Recognition, 2024, 148
- [6] Bao Xiaomin, WANG Jingbo. IoT technology in agricultural pests telemetry applications and development [J]. Journal of Xinjiang Agricultural Mechanization, 2020, (5): 25-29. DOI: 10.13620 / j.carol carroll nki issn1007-7782.2020.05.007.
- [7] Wu Lin, SUN Jingyu. Multi-branch RA capsule Network and its application in image classification [J]. Computer Science, 2022, 49 (06): 224-230.
- [8] Qiao Yu-si, Qiao Lian-Jun. Analysis of the Development Status and Industry Prospect of Embedded System [J]. Statistics and Management, 2013, (03): 89-90.
- [9] Duan Gui-duo, Luo Guang-chun, Zhang Li-Zong, zhu Dayong, Wang Zi-peng. Fundus Retinal image classification method Based on capsule Network [P]. China: CN201811197008.6.
- [10] Chen Jian, ZHOU Ping. Chinese Handwriting Identification Algorithm Based on Capsule Network [J]. Journal of Packaging, 2018, 10 (05): 51-56.
- [11] Feng Fenling, YAN Meihao, LIU Chengguang, et al. Export demand forecasting of China-Europe Train based on IPSO-Capsule-NN model [J]. China Railway Science, 2020, 41 (02): 147-156.
- [12] Sheng Guanqun, FANG Hao. Method and system for picking up microseismic P-wave based on capsule neural network [P]. China :CN202010057384.6.
- [13] Hu Chuntao, Xia Lingling, Zhang Liang, et al. Comparison of Text Classification Based on Capsule Network and Convolutional Network [J]. Computer Technology and Development, 2020, 30 (10): 86-91.
- [14] Huang Xiaogang, Huang Runcai, WANG Guijiang, et al. Facial expression Recognition Based on Feature Extraction and Capsule Network [J]. Intelligent Computer & Application, 2022, 12 (10): 68-74.
- [15] Geng Shaoguang. Design and Research of intelligent Car based on embedded device and deep learning Model [J]. Digital Technology and Application, 2020, 38 (05): 161-162. DOI:10.19695/j.cnki.cn12-1369.2020.05.90.
- [16] Application analysis of agricultural machinery automation and agricultural intelligence [J]. Agricultural Engineering Technology, 2023, 43 (29): 31-32. (in Chinese) DOI:10.16815/j.cnki.11-5436/s.2023.29.011.