# RESEARCH ON FRUIT IMAGE PROCESSING CLASSIFICANTION AND RECOGNITION BASED ON RESNET50 NEURAL NETWORK

**Qinghui Xu[1,a,*],Chen Luo [2],Dou Wen[3] ,Runhua Yang[4], Yilin Wan[2],Zongqi Ge[2]**

[1]*Faculty of Data Science and Information Technology, INTI International University, Nilai 71800, Negeri Sembilan, Malaysia*

[2]*College of Computer Science and Technology,University of East London,Longbridge Road, Dagenham, Essex RM8 2AS, UK*

[3]*Southwest Jiaotong University Hope College. No.8 Xuefu Avenue, Jintang County, Chengdu City, Sichuan Province, 610400, China.*

[4]*Yunnan Key Laboratory of Service Computing, Yunnan University of Finance and Economics, Kunming, Yunnan 650221, PR China*

[a]*qinghui.xu.sg@gmail.com,*

**corresponding author,*

***Abstract:****Convolutional Neural Network (CNN) is the most widely used algorithm model in computer vision problems , and has achieved remarkable results. In this paper, for the Fruits 360 image dataset on Kaggle, we use the algorithm based on The pre-trained network of Resnet-50, constructs the algorithm to build a convolutional neural network model. The difference between ResNet and traditional convolutional neural networks is that it adds some shortcuts to the network, i.e. direct channels are able to skip certain layers to optimize the network.ResNet introduces the idea of residual function F(x), which is set as: H(x)=F(x)+x. Specifically, if F(x)=0, then the residual functions form an identity map H(x)=x. It is easier for a neural network to fit the residuals.*

***Key words:*** *Convolutional Neural Network (CNN); ResNet; Residuals*

**1. Introduction**

Convolutional Neural Network is a typical structure of Artificial Neural Networks(ANN)[1], which belongs to feedforward neural network and is usually used to process image type data. Due to its own advantages,CNN can achieve the purpose of hierarchical abstraction of input image representation through specific local operations[2]. The successful application of CNN architecture is mainly due to its two advantages:first,CNN retains the binary space characteristics of the input image, and the correlation between the pixels of two-dimensional images is relatively high. CNN can realize sparse connection, which greatly reduces the number of parameters. Second, CNN can realize feature sharing, and the same position of each channel of the feature map is the result of the same convolution kernel after convolution calculation[3].

The process of forward propagation of the convolutional neural network to calculate the output is the feedforward process[4]. This process only propagates forward to calculate the output, and the output of the previous layer is used as the input of the next layer until the output of the last layer is obtained, and the network parameters are not adjusted in the process of forward propagation. In the process of Back Propagation of neural network, the error is propagated forward from the last layer through the calculation of the loss function, and the parameters of each layer are modified to adjust the weight parameters of the whole network[6].

Essentially, CNN can map the input to the output. Given the input and loss function, the convolutional neural network can automatically update the weights after building the network structure. Similar to ANN, CNN is also a multi-layer structure, each convolution layer can be seen as a number of two-dimensional convolution kernels, each convolution kernel is composed of multiple neurons. Each convolutional layer in CNN usually has n convolution kernels, so that the feature maps can share weights, which greatly reduces the number of parameters and speeds up the operation compared with the fully connected neural network[7]. A typical CNN structure consists of a feature extractor (filter)[8] and a pooling layer. There are two types of pooling layers, mean pooling and max pooling. The Pooling operation is similar to the convolution operation in that a specific operation is performed on the data in the partitioned region[9].

## 2. Data processing.

### *2.1Dataset characteristics.*

For the research topic CNN-based fruit image recognition, we used a kaggle competition: Fruits 360, where the training dataset contains 90,380 images of 131 fruits and vegetables[10]. Based on the fruit dataset, this study constructs a convolutional neural network based on the resnet-50 pre-trained model for classification prediction[11].

### *2.2Dataset preprocessing.*

One problem in image classification is class imbalance, where the number of samples per class varies greatly. I then create new images by changing existing images by rotating and flipping, here I use the Augmentor package to create new versions of images in a specified category and save them to a new directory, thus enlarging smaller classes[12]. So I started with class balance. First, I reduced the number of samples in the larger class by randomly removing samples[13]. Second, I increased the number of samples in the smaller class to achieve class balance

## 3. Construct the classification model.

### *3.1.Building a Convolutional Neural Network model.*

CNN is mainly composed of Convolutional Layer, Pooling Layer and Fully connected layer[14]. A convolutional layer extracts features using a convolution kernel and determines the position relationship between the extracted features and other features. For each convolution kernel, the input of a neuron is connected to the local receptive field output of the previous layer. Usually a convolutional layer is followed by a pooling layer for downsampling[15]. The structure of a convolutional neural network is shown in FIG. 1.

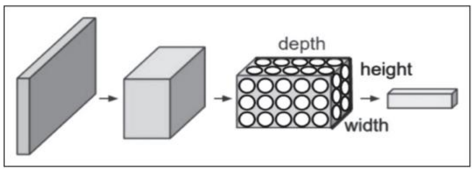


FIG. 1 Convolutional neural network architecture

The original image data is sent to the convolutional neural network after resize into a uniform size[16]. The convolutional layer can take a local unit (local receptive field) output of the previous layer as input for calculation each time, and perform convolution operation according to the set step size until the calculation of the entire input feature map is completed, realizing weight sharing[17]. The lower level kernels can extract some low-level features, such as edges, textures, etc. Higher-level kernels can extract higher-level features, such as complete local features, patterns, contours, etc. A convolutional layer usually contains n convolutional kernels[18]. The pooling layer usually follows the convolutional layer to downsample the data and increase the tolerance of the image to translation and deformation[19]. In the complete CNN network, the convolutional layer and the pooling layer present the state of alternating distribution.

***3.2.ResNet structure diagram.***

The residual network improves on the idea of cross-layer links in the highway network: using "shortcut connections", the input x is passed directly to the output as the initial result, and the output is: H (x) = F (x) + x. If F (x) = 0, then H (x) = x, which is the identity map mentioned earlier[20]. On this basis, the ResNet learning target is changed, instead of learning the complete feature output through layers of neural networks, the difference between the target value H (x) and x is learned, and the residual is: F (x) = H (x) - x. In this paper, the ResNet-50 network structure is used to overcome the shortcomings of inaccurate recognition due to the change of image chromaticity caused by the change of fruit ripening[21].

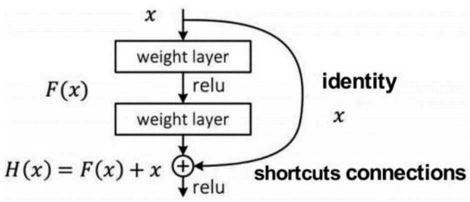


FIG. 2 ResNet residual unit

***3.3.ResNet50 network model.***

ResNet50 contains 49 convolutional layers and 1 fully connected layer. Among them, IDBLOCKx2 in the second to fifth stages represents two residual blocks without changing the size, CONVBLOCK represents the residual block with adding scale, and each residual block contains three convolutional layers[22]. So there are 1+3x(3+4+6+3)=49 convolutional layers and the structure is shown in Figure 3.

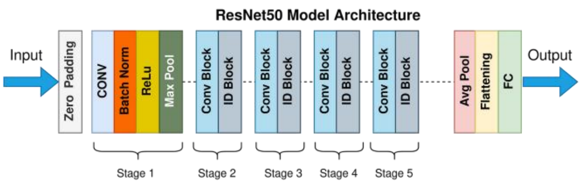


FIG. 3 ResNet50 structure diagram

CONV in FIG. 3 is the convolutional layer of the convolution operation, BatchNorm is the batch regularization processing, Relu is the activation function, MAXPOOL represents the maximum pooling operation, AVGPOOL represents the global average pooling layer operation, and stage1 to stage5 represents the residual block[23]. The size of the input data is 256x256x3. Since the ResNet50 neural network input data size is 224x224x3, image preprocessing needs to be carried out before the input data, the size of the non-standard data is cropped to the specified size of the data, and the normalization process is carried out[24]. Each channel average is subtracted from the channel average of the training set. After the continuous convolution operation of the residual block, the number of channels of the image pixel Matrix is getting deeper and deeper, and then the size of the image pixel matrix is changed by the Flatten layer. Finally, the image pixel matrix size is input to the full connection layer FC[25].

***3.4.SGD Classifier Classifier and confusion function.***

You know three kinds of classifiers, which are the Multinomial Naive Bayes classification algorithm, the linear support vector Machine classification algorithm, and the stochastic gradient descent classification algorithm SGD Classifier, which is a Classifier that quickly solves a classification algorithm with a loss function form[26]. Therefore, it can do linear regression, logistic regression, linear support vector machine.

Stochastic gradient descent classification algorithms fall into three categories according to how training data is used to calculate the training error: Batch gradient descent, Stochastic gradient descent, SGD, Mini-Batch gradient descent (MBGD)[27].

***2.4.1.Batch gradient descent.*** We use all the samples in each iteration, which has the advantage that we're looking at all the samples in each iteration, so we're looking at global optimization. Note that the name is not exact, but it is commonly used in the machine learning community[28]. Its disadvantage is that the training error of all samples in the training set has to be calculated in each iteration, which is not efficient when the amount of data is large[29].

***2.4.2Stochastic gradient descent.*** Each iteration randomly draws one sample from the training set, and in extremely large sample sizes, it may be possible to obtain a model with an acceptable loss without having to sample all the samples. The disadvantage is that a single sample may bring noise, resulting in not each iteration towards the overall optimal direction[30].

***2.4.3Mini-batch gradient descent.*** It is between batch gradient descent and stochastic gradient descent. Each iteration randomly draws a certain amount of data from the training set for training[31].

***3.5.The mixing matrix.***

Confusion matrix is in addition to the ROC (Receiver Operating Characteristic) curve and AUC (Area Under Curve) outside the degree of another judge classification[32].

| Real results | Prediction results | |
|---|---|---|
| | Positive example | Counter example |
| Positive example | TP(True Positive) | FN(False Negative) |
| Counter example | FP(False Positive) | TN(True Negative) |

Here are a few concepts to start with:

TP (True Positive) : The true value is 0 and the prediction is 0

FN (False Negative) : True is 0, predicted is 1

FP (False Positive) : True is 1, predicted is 0

TN (True Negative) : The true value is 1, and the prediction is also 1

precision=$\frac{\mathrm{TP}}{\mathrm{TP+FP}}$

The accuracy of the prediction is 0

sensitivity=recall=$\frac{TP}{TP+FN}$

Accuracy where the true is 0

f1score=2*precision*recall/precision+recall

## 4. Experiments.

### *4.1.Experimental platform and data set.*

The network platform used in the experiment is TensorFlow, which is a deep learning framework based on Python language and a symbolic mathematics system based on dataflow programming. It is widely used in the programming implementation of various machine learning algorithms[33]. It uses the language python and builds the OpenCV and TensorFlow environments. The dataset uses the fruit dataset provided by kaggle, which contains 90,380 images of 131 fruits and vegetables[34].

### *4.2.Experimental Ideas.*

The amount of data in this experiment is very large and the purpose of the experiment is to realize image classification. First, we think of the Convolutional Neural Network (CNN) model, first import the tensorflow package, Augmentor package, and split the data set according to the test, validation and test data[35]. The number of images of each fruit class in the training dataset was found that different data types directly had different amounts of data and the gap in the amount of data was very large[36]. I decided to experiment with the method of balancing the amount of data of each class to remove random examples from the category with the most data[37]. Saving them to a new directory to achieve class balance, I then merged the training dataset with the generated image dataset[38]. I used a linear classifier as a baseline, load the images using OpenCV, resize and normalize them, and then save them flat and vectorized in the set array with the appropriate labels[39].

The Stochastic Gradient Descent (SGD) [40]classifier is employed as the optimization method for a linear classifier, capable of addressing both binary and multiclass classification tasks. SGDClassifier[41] offers several hyperparameters that can be fine-tuned to optimize performance for specific tasks[42]. One notable advantage of the SGDClassifier[43], in comparison to the Support Vector Classifier (SVC)[44], is its significantly faster performance when handling large-scale training datasets[45].

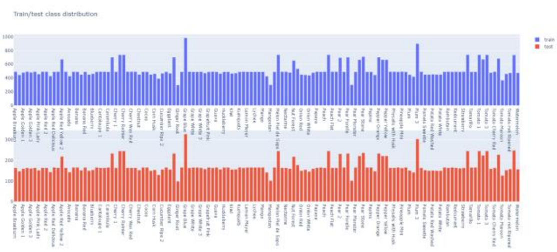

FIG. 4 ResNet50 Training set and test set

### *4.3.Experimental Modeling.*

Common algorithms in Convolutional Neural Networks (CNNS)[46] are AlexNet, VGG, Inception, ResNet, and the reason why I chose ResNet50: It is a pre-trained Convolutional Neural Network introduced by Microsoft Research in 2024[47]. It is a deep residual network capable of efficiently training deep neural networks with hundreds of layers. ResNet50[48] has 50 layers and uses skip connections to add the raw input of the layer blocks to its output. These connections make it easier to train deeper networks and also help alleviate the problem of vanishing gradients[49]. ResNet50 achieves state-of-the-art performance on a range of computer vision tasks, including image classification, object detection, and semantic segmentation[50]. A pre-trained ResNet50 neural network was employed to develop a deep learning model for feature extraction. Fully connected layers were subsequently added to the model to classify the images into 131 distinct classes[51].

***4.4.Analysis of model characteristics.***

*4.4.1.Efficiency of feature extraction*

For artificial intelligence tasks, the biggest difference between traditional machine learning and deep learning is that traditional machine learning needs to manually build some powerful features to achieve good results[52]. One of the biggest characteristics of deep learning is that we do not need to think about designing features, but use the deep structure of the network to help us automatically extract strong features, and the model will automatically capture the interdependence between features[53]. This not only greatly reduces the complexity of artificially designed features, but also achieves good generalization performance[54].

*4.4.2.Simplicity of data format*

In traditional machine learning, we need to do a lot of preprocessing on the original data set, compare data normalization, format conversion, etc., but in neural network, we do not need to do too much preprocessing on the data, even if there are some wrong labels in the data, the network can also show good adaptability[55].

***4.5.Analysis of experimental results.***

In multi-class classification tasks, it is not suitable to use PR curve and ROC curve for metric evaluation, but we can still use confusion matrix to deal with it. You can use the conf_matrix () function to determine whether the classification is good or bad; Figure 5 shows the result of the confusion matrix[56].

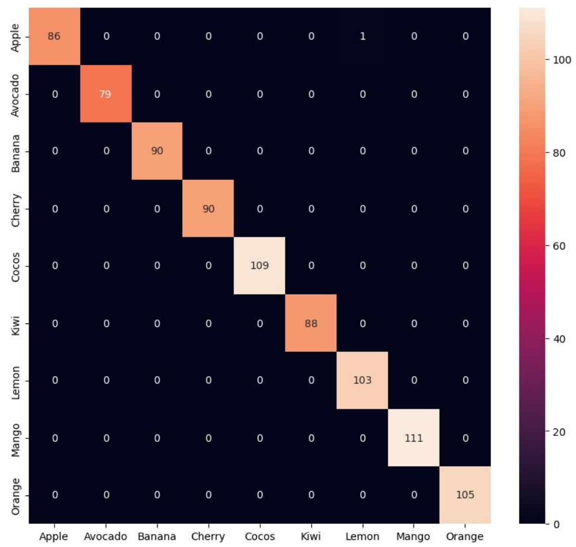


FIG. 5 Confusion Matrix

After training for 25 epochs, the accuracy of the ResNet50 model on the training set is over 97%, and the accuracy on the validation set is 98%[57]. By the time the model is trained for 25 epochs, the error on the training set is almost zero, and the error on the validation set is also below 0.2, and the average error on the validation set is 0.28. It shows that the model has fast convergence speed, high recognition accuracy, high accuracy on the validation set, and low error ratio[58]. Moreover, the error of the model on the validation set is lower than that of the trainer, which indicates that the overfitting is well suppressed and the image recognition effect is better. The accuracy and error of the ResNet50 model on the training set and the validation set are shown in FIG. 6[59].

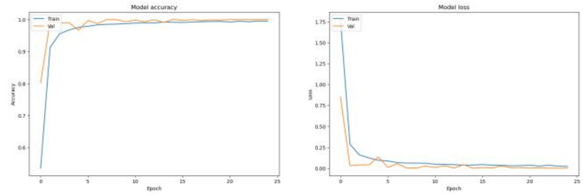

FIG.6 Model accuracy and Model loss

**5. Conclusion.**

In this paper, combined with the latest technology in development, a deep learning network model based on resnet50 is proposed to realize high-precision recognition and classification of fruit images. The residual block in the model effectively overcomes the problems such as network degradation and ensures that the performance will not degrade while the depth of the network is deepened. In view of the large and complex fruit image dataset problem, this paper uses the means of transfer learning to avoid the appearance of overfitting problem. This model has the advantages of greater depth, faster convergence, higher accuracy and easy generalization. It has made many feasibility demonstrations in the application of computer artificial neural network and image processing classification to life, which provides effective guidance for the development of computer vision technology and artificial intelligence.

**6. References.**


*[1] Hindarto D. Use of RESNET-50 Neural Network in Diagnosing Diseases Mango Leaves[J]. Journal of Computer Networks, Architecture and High Performance Computing, 2024, 6(1): 220-230.*

*[2] Lin H, Yu L, Wan M, et al. Fruit image classification based on ResNet50 and transfer learning[C]//Fourth International Conference on Computer Vision and Data Mining (ICCVDM 2023). SPIE, 2024, 13063: 1306302.*

*[3] Upadhyay L, Saxena A. Evaluation of Enhanced Resnet-50 Based Deep Learning Classifier for Tomato Leaf Disease Detection and Classification[J]. Journal of Electrical Systems, 2024, 20(3s): 2270-2282.*

*[4] Prabha R, Selvan K. Modified RESNET50 with attention module for detection and classification of pests in vegetable crops[J]. Journal of Advanced Research in Applied Sciences and Engineering Technology, 2024: 67-86.*

*[5] Sinha B B, Dhanalakshmi R. A multi-fused convolutional neural network model for fruit image classification[J]. International Journal of Cognitive Computing in Engineering, 2024, 5: 416-424.*

*[6] Salem D A, Hassan N A A, Hamdy R M. Impact of transfer learning compared to convolutional neural networks on fruit detection[J]. Journal of Intelligent & Fuzzy Systems, 2024 (Preprint): 1-13.*

*[7] Kumar M, Arora A, Deb A, et al. Deep Learning for Accurate Plant Disease Classification Using ResNet50: A Comprehensive Approach[C]//2024 International Conference on Computational Intelligence and Computing Applications (ICCICA). IEEE, 2024, 1: 125-130.*

*[8] Li T, Zhu W, Che X. Design and Implementation of Apple Leaf Disease Recognition System Based on ResNet50[J]. Journal of Computing and Electronic Information Management, 2024, 12(2): 96-104.*

*[9] Fadhlurrahman M F, Wirayuda T A B, Purnama B. Prediction of Fruit Freshness Level Using*

*Convolutional Neural Network[C]//2024 12th International Conference on Information and Communication Technology (ICoICT). IEEE, 2024: 137-143.*

*[10] Toptaş B, Güven S A. Guava fruit classification system design with convolutional neural networks[J]. Gümüşhane Üniversitesi Fen Bilimleri Dergisi, 2024, 14(4): 1247-1258.*

*[11] Raza S M, Raza A, Babeker M I A, et al. Efficient citrus fruit image classification via a hybrid hierarchical CNN and transfer learning framework[J]. Journal of Food Measurement and Characterization, 2024: 1-22.*

*[12] Khan S, Yadav A K, Shukla S S P, et al. ResNet-50: A breakthrough in plant disease detection accuracy for agricultural empowerment[M]//Computational Methods in Science and Technology. CRC Press, 2024: 364-371.*

*[13] Vo H T, Mui K C, Thien N N. Fusing Xception and ResNet50 features for robust grape leaf disease classification[J]. Bulletin of Electrical Engineering and Informatics, 2025, 14(1): 268-275.*

*[14] Sönmez D P, Kılıç Ş. Deep Learning for Automatic Classification of Fruits and Vegetables: Evaluation from the Perspectives of Efficiency and Accuracy[J]. Türkiye Teknoloji ve Uygulamalı Bilimler Dergisi, 5(2): 151-171.*

*[15] Kunduracioglu I. Utilizing ResNet Architectures for Identification of Tomato Diseases[J]. Journal of Intelligent Decision Making and Information Science, 2024, 1: 104-119.*

*[16] Bai J, Mei Y, He F, et al. Rapid and non-destructive quality grade assessment of Hanyuan Zanthoxylum bungeanum fruit using a smartphone application integrating computer vision systems and convolutional neural networks[J]. Food Control, 2025, 168: 110844.*

*[17] Song X, Mariano V Y. Fruit Tree Disease Recognition Based on Residual Neural Network and Attention Mechanism[J]. Journal of Artificial Intelligence and Technology, 2024, 4(2): 145-152.*

*[18] Badriyah A, Sarosa M, Asmara R A, et al. Grapevine Disease Identification Using Resnet−50[C]//BIO Web of Conferences. EDP Sciences, 2024, 117: 01046.*

*[19] Kumaar D M, Palani S. ResNet50 Integrated Vision Transformer for Enhanced Plant Disease Classification[C]//2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT). IEEE, 2024: 1-6.*

*[20] Soni T. Deep Learning for Accurate Grape Disease Classification with ResNet50[C]//2024 5th International Conference on Smart Electronics and Communication (ICOSEC). IEEE, 2024: 816-820.*

*[21] Ratha A K, Behera S K, Devi A G, et al. Optimizing precision agriculture: Bayesian-enhanced papaya (Carica papaya L.) fruit disease classification via cubic SVM and ResNet-101 deep features[J]. Journal of Intelligent & Fuzzy Systems (Preprint): 1-17.*

*[22] Khalifa N E M, Wang J, Hamed N. Taha M, et al. DeepDate: A deep fusion model based on whale optimization and artificial neural network for Arabian date classification[J]. Plos one, 2024, 19(7): e0305292.*

*[23] Xiao B, Nguyen M, Yan W Q. Fruit ripeness identification using YOLOv8 model[J]. Multimedia Tools and Applications, 2024, 83(9): 28039-28056.*

*[24] Vera O, Cruz J, Huaquipaco S, et al. Automated and Enhanced Classification of Persea Americana using optimized Deep Convolutional Neural Networks with improved training strategies for Agro-Industrial Settings[J]. IEEE Access, 2024.*

*[25] Nasra P, Gupta S. ResNet50: Deep Learning Method for Automated Detection of Banana Leaf Spot Diseases[C]//2024 First International Conference on Innovations in Communications, Electrical and Computer Engineering (ICICEC). IEEE, 2024: 1-6.*

*[26] Karki S, Basak J K, Tamrakar N, et al. Strawberry disease detection using transfer learning of deep convolutional neural networks[J]. Scientia Horticulturae, 2024, 332: 113241.*

*[27] Zhu Y, Chen S, Yin H, et al. Classification of oolong tea varieties based on computer vision and convolutional neural networks[J]. Journal of the Science of Food and Agriculture, 2024, 104(3): 1630-1637.*

*[28] Mazurets O, Molchanova M, Klimenko V, et al. Datalogic Model for Image Recognition by Convolutional Neural Network Using Cloud Services[J]. 2024.*

*[29] Pajany M, Venkatraman S, Sakthi U, et al. Optimal Fuzzy Deep Neural Networks based Plant Disease Detection and Classification on UAV-based Remote Sensed Data[J]. IEEE Access, 2024.*

*[30] Sari I P, Warsito B, Nurhayati O D. Implementation of Convolutional Neural Network for Classification of Helminths Egg Based on Image Processing[C]//2024 International Conference on Electrical Engineering and Computer Science (ICECOS). IEEE, 2024: 38-43.*

*[31] Sari I P, Warsito B, Nurhayati O D. Implementation of Convolutional Neural Network for Classification of Helminths Egg Based on Image Processing[C]//2024 International Conference on Electrical Engineering and Computer Science (ICECOS). IEEE, 2024: 38-43.*

*[32] Arivalagan D, Nikitha P, Manoj G, et al. Intelligent Fruit Quality Assessment Using CNN Transfer Learning Techniques[C]//2024 International Conference on Distributed Computing and Optimization Techniques (ICDCOT). IEEE, 2024: 1-6.*

*[33] Saedi S I, Rezaei M, Khosravi H. Dual-path lightweight convolutional neural network for automatic sorting of olive fruit based on cultivar and maturity[J]. Postharvest Biology and Technology, 2024, 216: 113054.*

*[34] Rybacki P, Niemann J, Derouiche S, et al. Convolutional Neural Network (CNN) Model for the Classification of Varieties of Date Palm Fruits (Phoenix dactylifera L.)[J]. Sensors, 2024, 24(2): 558.*

*[35] Wang N, Yu S K, Qi Z P, et al. Pears classification by identifying internal defects based on X-ray images and neural networks[J]. Advances in Manufacturing, 2024: 1-10.*

*[36] Chen H, Zhou G, He W, et al. Classification and identification of agricultural products based on improved MobileNetV2[J]. Scientific Reports, 2024, 14(1): 3454.*

*[37] Chen H, Zhou G, He W, et al. Classification and identification of agricultural products based on improved MobileNetV2[J]. Scientific Reports, 2024, 14(1): 3454.*

*[38] Özbay E, Özbay F A, Gharehchopogh F S. Visualization and classification of mushroom species with multi-feature fusion of metaheuristics-based convolutional neural network model[J]. Applied Soft Computing, 2024, 164: 111936.*

*[39] Kusakunniran W, Imaromkul T, Aukkapinyo K, et al. Automatic classification of mangosteens and ripe status in images using deep learning based approaches[J]. Multimedia Tools and Applications, 2024, 83(16): 48275-48290.*

*[40] Cao R. Research on The Efficiency of Mixed Commodities Classification based on Deep Convolutional Neural Networks[J]. Highlights in Science, Engineering and Technology, 2024, 85: 997-1004.*

*[41] Cao R. Research on The Efficiency of Mixed Commodities Classification based on Deep Convolutional Neural Networks[J]. Highlights in Science, Engineering and Technology, 2024, 85: 997-1004.*