



# Research on identification method of tangerine peel year based on deep learning

Ziyi CHU<sup>1</sup> , Fengmei LI<sup>1</sup>, Dongwei WANG<sup>1</sup>, Shusheng XU<sup>1</sup>, Chunfeng GAO<sup>1</sup>, Haoran BAI<sup>1\*</sup>

## Abstract

Tangerine Peel has rich medicinal value, known as ‘one kilogram of tangerine peel, one kilogram of gold’. However, the value of tangerine peels in different years is different, and there is no significant difference in the appearance of tangerine peels in different years. Identifying their authenticity has brought trouble to the industry. Generally speaking, the characteristics of tangerine peel can be identified through the texture, color and oil parcel points on the surface of tangerine peel. However, compared with the feature recognition of other Chinese medicinal materials, there is no significant difference in the shape of tangerine peel in different years, and the color is similar. Therefore, the feature extraction of tangerine peel is more complicated and the recognition is more difficult. The existing deep learning algorithms face great challenges in efficient and high accuracy recognition. In response to this challenge, this paper builds a new lightweight tangerine peel recognition algorithm TPRA (Tangerine Peel Recognition Algorithm) based on ResNet50. This algorithm uses a variety of methods to optimize the generalization ability of the model and improve the recognition accuracy. Firstly, TPRA adopts mixed data enhancement, including traditional data enhancement, deep convolution generation confrontation network DCGAN, and Mosaic data enhancement to enhance the richness of sample images in the dataset, reduced the data of each batch regularization (Batch Normal), and enhanced the performance of algorithm identification. Secondly, TPRA introduced the attention mechanism module CBAM (Convolutional Block Attention Module) combined with the cross stage partial network CSPNet (Cross Stage Partial Network) to propose an improved ResNet50 model, which adjusts the position of the maximum pooling layer and disassembles the large convolution kernel to effectively avoid overfitting. The experimental results showed that the accuracy of the algorithm can reach 98.8%, and the effect was better than that of Alexnet, VGG16 and Resnet50. TPRA provided a new method for the identification of peel years.

**Keywords:** image recognition; Mosaic; resnet50; DCGAN; CSPNet.

**Practical Application:** Non-destructive classification of tangerine peel.

## 1 Introduction

Tangerine Peel is the dried ripe peel of Rutaceae tangerine and its cultivated varieties, which can be directly used for clinical treatment (Yi et al., 2015). Tangerine Peel can inhibit respiratory inflammation (bronchitis and asthma, etc.) (Ho & Kuo, 2014). The contents of organic compounds in pericarps of different years are also significantly different, and the longer the storage year is, the higher the total flavonoids content is (Lin et al., 2008). Because of its important medicinal value in recent years, the image recognition of traditional Chinese medicine represented by tangerine peel has attracted wide attention of scholars. Deep learning is a machine learning technology used to establish and simulate the neural network of human brain for analysis and learning, and imitate the mechanism of human brain to explain data. The most prominent application is the field of computer vision (Zhang & Xu, 2022). Computer vision is based on the main features of images. Each image has its own characteristics. For example, in the tangerine peel dataset, different years of tangerine peel contain different image features. Computer vision uses these features to learn and establish models, so as to achieve the effect of tangerine peel year recognition. In the process of tangerine peel year recognition, aiming at the problem

of insufficient samples and unsatisfactory feature extraction of data sets, this paper introduces deep convolution generation adversarial network (DCGAN) (Radford et al., 2015) and Mosaic data enhancement to further enhance the data set, expand the sample size of the data set and strengthen the recognition of local features. By comprehensively comparing the accuracy, reasoning time, power consumption, computation and other data of deep learning algorithms, it is found that ResNet (He et al., 2016) has better comprehensive performance. The residual network solves the problem of network degradation and avoids the problem of gradient disappearance caused by too deep network. But there are problems of long training time, large number of model parameters and large power consumption. Therefore, this study improves the accuracy of algorithm identification by adjusting the position of the maximum pooling layer, reducing the training time of the model and dismantling the large convolution kernel. It introduces the attention mechanism module CBAM combined with the CSPNet module to strengthen the learning ability of the model, strengthen the extraction of target features, and remove the calculation bottleneck layer to reduce the consumption of multiple pools and reduce the memory occupation.

Received 15 May, 2022

Accepted 05 July, 2022

<sup>1</sup>College of Mechanical and Electrical Engineering, Qingdao Agricultural University, Qingdao, China

\*Corresponding author: baihaoran111@126.com

## 2 Related work

The traditional year identification of dried tangerine peel often subjectively judges the appearance, color, oil bag and other characteristics of dried tangerine peel by manual, which requires high experience of staff, is easy to misjudge and has long recognition time. With the application of deep learning technology in the field of agriculture becoming more and more extensive, a large number of scholars in China and abroad have proposed relevant research methods for Chinese medicinal materials image recognition. For example, Gu et al. (2010) used the traditional convolution neural network to construct the image database, and combined with migration learning to solve the problem of intelligent identification of Chinese medicinal slices. Liu et al. (2018) trained 50 kinds of Chinese herbal medicine images collected under natural conditions through GoogLeNet to solve the problem of a wide variety of Chinese herbal plants and similar appearance of some Chinese herbal plants. A method for automatic identification and classification of Chinese herbal medicine by processing images and deep learning was designed. Dileep & Pournami (2019) uses Alexnet model to extract features from data sets and classifies them by SoftMax and SVM classifier, which has great practical value for classification and recognition of medicinal plants. Kan et al. (2017) classified 12 kinds of medicinal plant leaves by preprocessing leaf images combined with SVM classifier. Janani & Gopal (2013) trained artificial neural network (ANN) classifier to identify the leaves of medicinal plants, which is suitable for leaf recognition system with small input and short calculation time.

In recent years, the application of deep learning technology in agriculture, food, traditional Chinese medicine and other fields has developed rapidly (Chen & Yu, 2022), such as crop diseases and insect pests identification, fruit and vegetable varieties identification, and traditional Chinese medicine decoction pieces identification. Multi-task learning and attention mechanism are also widely used, and new algorithms with high accuracy have been proposed. For example, Hu et al. (2020) used the concept of multi-task learning, based on neural network and supplemented by traditional features, combined the two to establish a new deep learning model, and identified 30437 pieces of Chinese herbal medicine images. The accuracy rate reached 86.2%. Zou et al. (2022) used Optuna algorithm to optimize the XGBoost and LightGBM models respectively to realize the variety classification of peanut hyperspectral images. Xing et al. (2020) trained the images of Chinese herbal pieces based on DenseNet model and transfer learning, and the highest recognition rate can reach 97.34%. Xu et al. (2021) proposed a new recognition attention pyramid network for the construction of Chinese herbal medicine recognition framework. Mao et al. (2021) proposed a hierarchical structure of multi task visual perception based on R-CNN algorithm to realize food diet assessment. However, with the continuous improvement of algorithm accuracy, the amount of calculation is also increasing, and it is difficult to deploy to areas where hardware resources such as edge computing equipment, embedded equipment and mobile equipment are limited. To solve this problem and based on the former research, this paper proposes a hybrid data enhancement method for data enhancement. At the same time, this paper introduces the attention mechanism module CBAM

combined with CSPNet module to propose a new lightweight algorithm for tangerine peel year recognition.

## 3 Algorithm analysis

In the deep learning image recognition technology, the most important two parts are data set samples and model structure. The algorithm proposed in this paper is based on these two parts to make corresponding improvements for the recognition of tangerine peel years.

### 3.1 Data augmentation

The lack of data set samples will lead to low recognition accuracy and over-fitting. Therefore, the data enhancement of tangerine peel data set can effectively reduce over-fitting and improve accuracy.

#### *Traditional data enhancement*

Due to the insufficient number of original samples in the data set, data enhancement is performed on the data set to increase the sample size of the data set and improve the accuracy of the algorithm. Data enhancement can effectively reduce overfitting. Data enhancement is mainly divided into single sample data enhancement and multiple sample data enhancement. Single sample data enhancement increases the sample size of the data set by shearing, flipping, fuzzy algorithm, adding noise and other operations based on the existing original data set. To a certain extent, this method can reduce the problem of data imbalance and improve the generalization ability of the algorithm (Shijie et al., 2017). This manuscript mainly used single sample data enhancement method for tangerine peel year recognition dataset.

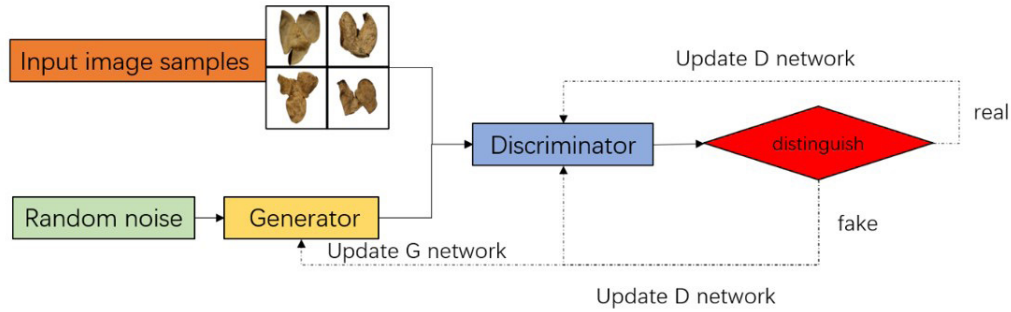
#### *Data enhancement based on DCGAN*

Generative Adversarial Network (GAN) (Goodfellow et al., 2020) was inspired by game theory by Goodfellow et al. (2020) Using the generative model learning method of confrontation thought. The basic frame diagram is shown in Figure 1.

The GAN network is composed of two parts: the generating network G (Generator) and the discriminant network D (Discriminator). G is the network for generating images, receiving random noise z, and generating images through this noise, denoted as G(z). D is the discriminant network to determine whether the image is true. The input parameter is that x represents an image, and the output D(x) represents the probability that x is a real image. G and D form a dynamic game process that achieves Nash equilibrium (Simonyan & Zisserman, 2014), expressed as (Equation 1):

$$\min_G \max_D V(D, G) = E_{x \sim P_{data}(x)} [\log D(x)] + E_{z \sim P_{z(z)}} [\log(1 - D(G(z)))] \quad (1)$$

x represents the real image, z represents the noise of the input G, G(z) represents the image generated by G, D(x) represents the probability of D discriminating the real image, and D(G(z)) is the probability of D discriminating the real image generated



**Figure 1.** GAN generates frame.

by G. For G, the larger the  $D(G(z))$  is, the better the  $V(D, G)$  becomes smaller. Therefore, for G, the minimum ( $\min_G$ ) is obtained. For D, the stronger the discriminant ability of D is, the greater  $D(x)$  is, and the smaller  $D(G(z))$  is, and then  $V(D, G)$  will become larger. So for D it is maximum ( $\max_D$ ). The process can be illustrated in Figure 2.

The G network in DCGAN is shown in Figure 3.

*Mosaic data enhancement*

Mosaic data enhancement (Bochkovskiy et al., 2020) is an improvement of CutMix. By cutting four sample images at random positions, each image has a corresponding image frame, and the cut four images are restitched to get a new sample image, as shown in Figure 4.

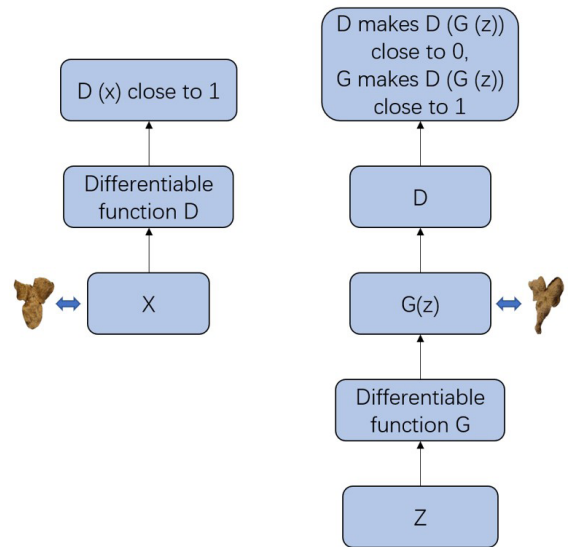
The image enhanced by mosaic data can exclude redundant information in the model training, improve the training efficiency, strengthen the recognition of local features, and integrate the information of multiple samples in the clipping region, which can better improve the recognition effect of the model. Compared with the traditional data enhancement technology, there is no problem of feature loss after image mixing and the addition of non-feature information, which can improve the performance of model classification. In the calculation of BN, four image data can be calculated at one time, which can reduce the memory occupation and shorten the training time.

**3.2 Model improvement**

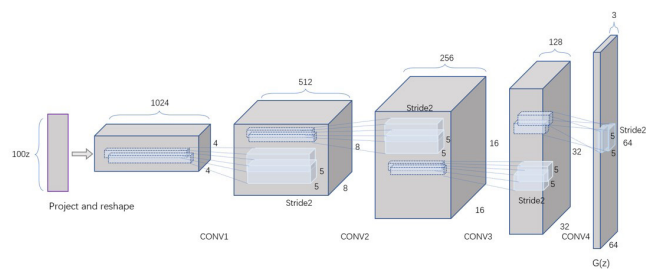
The change of the model structure will have different effects on the accuracy, and the improvement of the structure can be divided into the adjustment of the internal structure and the introduction of external corresponding modules. In this paper, aiming at the problem that the feature extraction of tangerine peel year recognition is more complex than that of other image recognition, the model is improved from the internal and external aspects.

*Introduction to ResNet50 model*

ResNet is conducive to solving complex tasks and improving detection accuracy. The main idea is to add direct channels in the network, allowing the original input information to be directly



**Figure 2.** Generating an adversarial network process diagram.



**Figure 3.** G network diagram.

transmitted to the following layer (Mukti & Biswas, 2019), as shown in Figure 5.

Through this method, the neural network at this layer can learn the residual of the output of the previous network instead of learning the whole output, so ResNet is also called residual network. ResNet50 has a total of 50 layers, and the 3-channel image of  $224 \times 224$  is input. After the first convolution, the 64-channel image of  $112 \times 112$  is processed, and then the 64-channel image

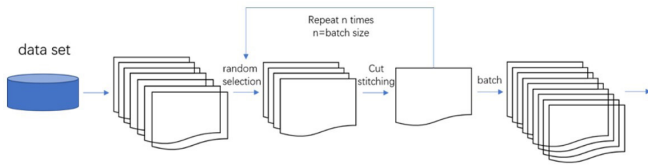


Figure 4. Mosaic process diagram.

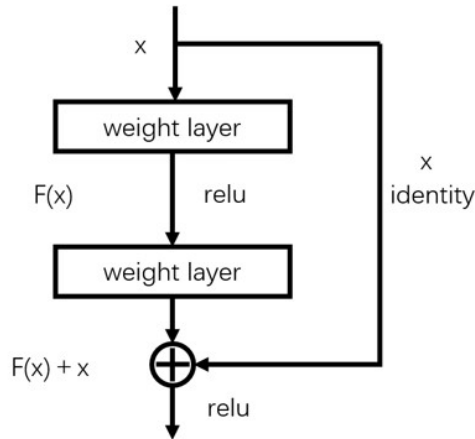


Figure 5. ResNet residual learning module.

of  $56 \times 56$  is transformed through the maximum pooling layer. After the processing of the three Bottleneck modules in the first stage, the dimension of the 256-channel image of  $56 \times 56$  is reduced to 1/2 of the previous stage, and the number of channels is twice of the previous stage. After the four stages, the 2048-channel image of  $7 \times 7$  is output. Finally, the model crosses the average pooling layer and full connection layer.

### Model optimization

Due to the increase of network layer and training data, ResNet50 model contained more structural features than other models, but there were problems such as too many model parameters, inaccurate target feature extraction in complex data sets and long training time. For this we used four ways to optimize.

#### 1. Adjust maximum pooling layer position

Change the location of the maximum pooling layer: exchange the location of the maximum pooling layer and the relu activation function. The pooling operation was first performed for down-sampling, and the consumption is reduced when the relu activation function was performed. The image features were compressed through the pooling layer, so that the image features become smaller and the amount of calculation was reduced. Then, the main features were extracted to accelerate the calculation speed and increase the receptive field. Therefore, the pooling efficiency was higher and the over-fitting was prevented

to a certain extent. After the exchange of sequence, the training time was reduced by about 5%, and the accuracy did not change. To further improve the training efficiency, the maximum pooling layer was moved to Batch Norm before pooling operation can shorten the training time. The training time was shortened by about 6.5%, and the training efficiency was greatly improved. The conv-pool block is shown in Figure 6. However, due to the slight influence of the change of the network structure on the test accuracy, we continued to optimize the next step.

#### 2. Dismantling large convolution kernels

In order to improve the accuracy of tangerine peel year identification and reduce the number of parameters, this paper replaced the large kernel convolution with the multi-layer small convolution kernel, which can not only reduce the parameters, but also increase the network depth to achieve large network capacity and complexity. The cascade of three  $3 \times 3$  convolutions was used to replace the convolution of  $7 \times 7$ , as shown in Figure 7.

Under the condition of ensuring the same receptive field, the model improves the depth of the network and the effect of the neural network. The images of data sets used in this paper were  $224 \times 224$ , and the convolution kernel of  $7 \times 7$  was used for convolution. The stride was 2, and the result was 218; the result using three convolution kernels of  $3 \times 3$  was that the first layer was 222, the second layer was 220, and the third layer was 218. The result was as same as the result of  $7 \times 7$  convolution. For three  $3 \times 3$  convolution kernels, the total number of parameters used was  $3 \times (3 \times 3) \times \text{Channels}$ . And for  $7 \times 7$  convolution kernels, the total number of parameters used was  $7 \times 7 \times \text{Channels}$ . So the number of parameters can be significantly reduced by about 40%.

#### 3. Introduction of convolutional attention module

CBAM (Convolutional Block Attention Module) represents the attention mechanism module of convolution module, which is a kind of attention mechanism module combining space and channel. Compared with Senet, it can achieve better results by focusing only on the attention mechanism of channel and ignoring the most important spatial attention part (Woo et al., 2018). In order to solve the problem that it was difficult to extract the main discriminant features of tangerine peel year identification, CBAM was added to the ResNet50 model to strengthen the extraction of target features from the two dimensions of space and channel. The output of the convolutional layer was used as the input of the CBAM module to enter the channel attention module for maximum pooling and average pooling. Then two features were obtained by MLP, and the output features were added. Finally, the weighted results were obtained through the sigmoid activation operation and enter the spatial attention module. The output of the channel attention module was subjected to maximum pooling and average pooling to obtain two 2D feature maps. The two feature maps were spliced, and the spatial attention feature map was obtained by convolution. Finally, the feature was adjusted with the input feature map to obtain the final output of the entire module, as shown in Figure 8.

However, after experiments in this model, adding this module to the convolution layer in each block will change the

structure of ResNet50 network, resulting in an increase in the number of parameters, unable to use pretraining parameters and the extraction of target features cannot achieve ideal results.

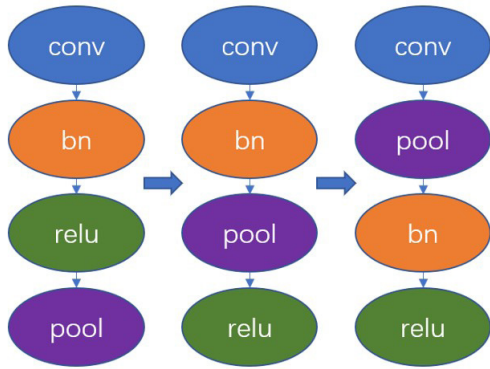


Figure 6. Position adjustment of maximum pooling layer.

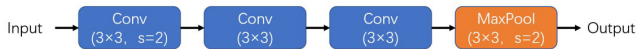


Figure 7. Model structure adjustment.

Therefore, this paper chooses to introduce CBAM after each block. Grad-CAM (Gradient-weighted Class Activation Map) is used to visually compare the extraction effects of ResNet50 and ResNet50 + CBAM for model feature extraction, as shown in Figure 9. In ResNet50, the extraction discriminant area of feature will deviate, while the feature extraction after CBAM is introduced focuses on the areas with obvious texture and oil parcel.

4. Combined with CSP module

CSPNet was proposed by Wang et al. (2020) in 2020. CSPNet divides each block into two parts according to the channel dimension of the feature map. One part is carried out according to the original process, the other part is directly connected to the block output, and then the two parts are merged through some transition layers.

In this paper, CSPNet module is combined with ResNet model. CSPNet maps the input feature map into two parts, one part carries out the original ResBlock operation, and the other part directly outputs concatenated with Partial ResBlock. Because the input channel is reduced by half, CSPNet does not need to introduce the bottleneck layer. CSPNet module is used to replace the original convolution operation, and the repeated information in the process of feature propagation is reduced by

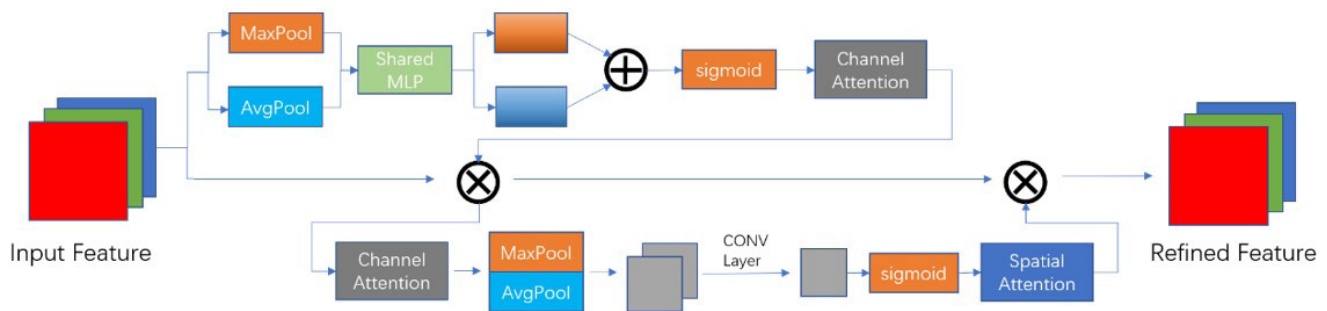


Figure 8. CBAM module structure diagram.

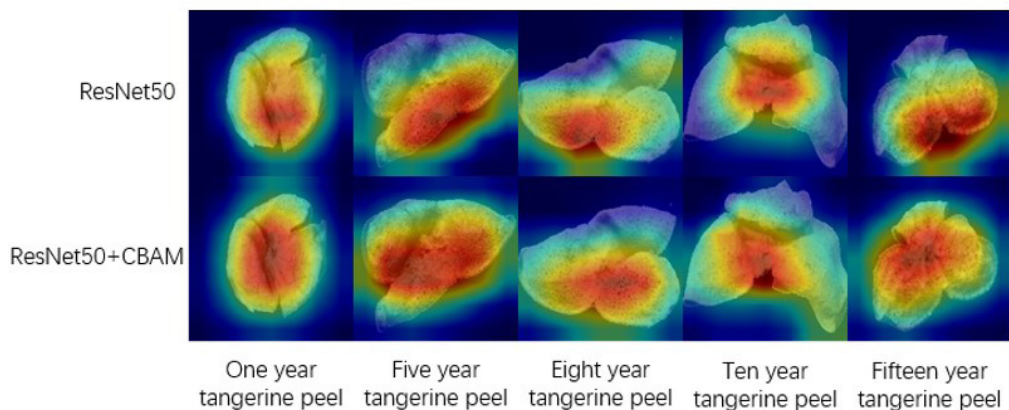


Figure 9. Effect comparison diagram.

cross-stage splitting and cascade strategy. The base feature map of a res block in ResNet is  $w \times h \times c$ , growing at  $d$ , with a total of  $z$  res layers. Then, the CIO of res block is  $(c \times z) + ((z^2 + z) \times d) / 2$ , and the CIO of partial res block is  $(c \times z) + (z^2 + z) \times d / 2$ . Partial res block can save up to half of the memory usage and improve the network operation speed. As shown in Figure 10 (Wang et al., 2020), it achieves better recognition effect and reduces the computational cost of the model. Because the gradient descent effect of the mish activation function is better, the mish activation function (Mish, 2019) is used to replace the relu activation function. The improved model structure is shown in Figure 11.

## 4 Experiment and result

### 4.1 Experimentation

#### Experimental environment

The main hardware and software used in this experiment are shown in Table 1.

#### Data preparation

The data set sample used in this study is Guangdong Xinhui tangerine peel. Because the classification of tangerine peel in different years is the same in appearance, it cannot be used as a classification feature. The main features are color, texture and oil package points, which have higher requirements for the data set. The industrial camera was used to take pictures of dried tangerine peel. A total of 613 dried tangerine peel images were collected and divided into five categories according to the experimental requirements. Each image contains a complete dried tangerine peel, including 121 dried tangerine peel images in one year, 118 dried tangerine peel images in five years, 136 dried tangerine peel images in eight years, 128 dried tangerine peel images in ten years and 110 dried tangerine peel images in fifteen years. These five types of images are stored in five different folders. Each folder represents a class of labels, and the sample image in the data set is uniformly processed into  $224 \times 224$  pixels. The image of the data set is shown in Figure 12.

#### Data augmentation

##### 1. Traditional data enhancement

Data set A is constructed by using the traditional data enhancement method. According to the characteristics of the

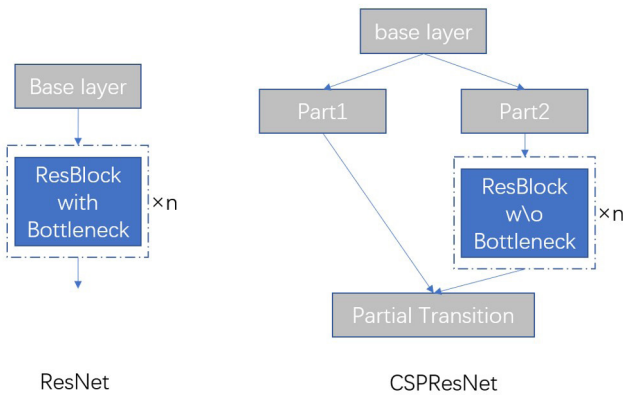


Figure 10. Application of CSPNet in ResNet.

Table 1. Operating environment-related parameters.

Name	Parameter
Memory	8GB
CPU	AMD Ryzen 7 4800H
GPU	NVIDIA GeForce GTX 1650
OS	Microsoft Windows 10 Flagship Edition (64-bit)
Software development environment	TensorFlow
Programming language	Python

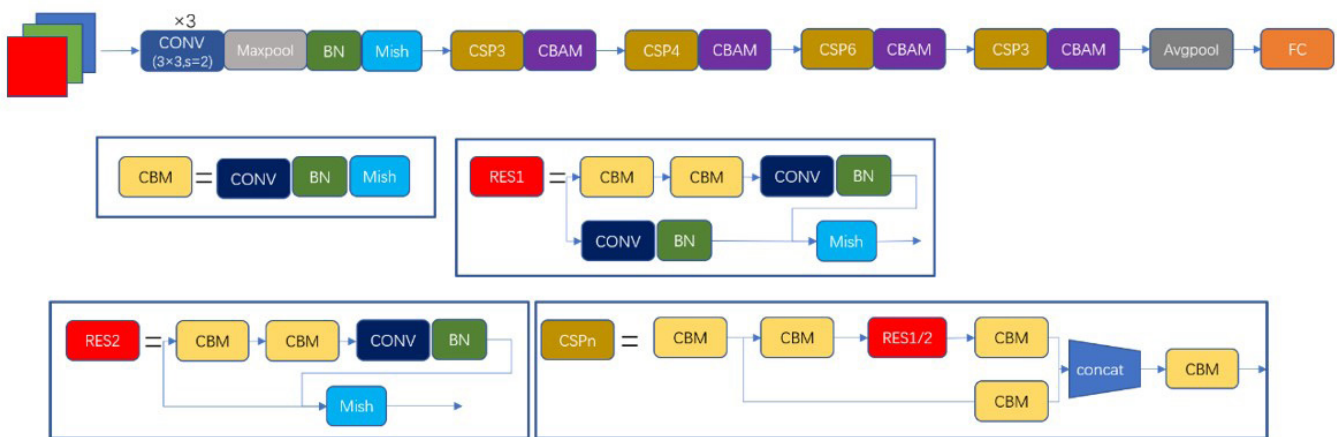


Figure 11. Improved model structure diagram.



**Figure 12.** Sample pictures of dried tangerine peels in different years.

classification of tangerine peel years, it is enhanced by four traditional data enhancement methods: horizontal and vertical flip, random clipping, translation rotation and adding noise. After data enhancement, the sample size of the data set increases from 613 to 5517, and the distribution of each year is shown in Table 2.

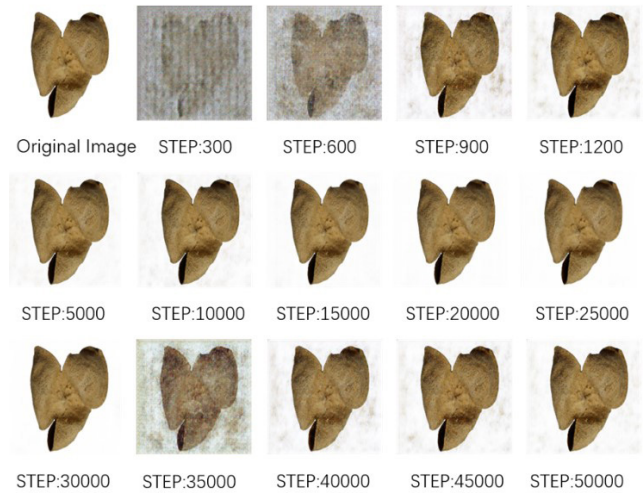
2. Data enhancement based on DCGAN

The data set B uses AMD Ryzen 7 4800 H and NVIDIA GTX1650, and the memory is 8G. GPU operation is performed on TensorFlow. Through multiple comparison experiments, it is found that the convergence effect is the best when the learning rate is 0.0002. Therefore, the learning rate is 0.0002, and 50000 epochs are performed. Stochastic Gradient Descent (SGD) (Bottou, 2012) and Adam optimizer (Kingma & Ba, 2014) are used to reduce the workload of parameter adjustment. The length of the vector sampled in the random distribution is set to 32, and the size of the input and output images is 224 × 224. The resulting image is shown in Figure 13. According to the experiment, the loss curve is drawn. As shown in Figure 14, d\_loss and g\_loss are in balance between 20000-34000 training times, and the generated images are more and more clear. After 34000 training times, d\_loss and g\_loss increase rapidly, and a large number of noise interference appears in the generated images. The sample size of the dataset after DCGAN data enhancement is shown in Table 3.

3. Mosaic data enhancement

Using python3.6 as the compiler and PIL, NumPy, matplotlib and other libraries, four samples four samples are randomly selected from the tangerine peel year data set each time. The four images are cut, flipped, scaled, and mirrored, respectively. The positions of the four segmented images are corresponding to each other, and the images and image frames are combined. The four sample images are cut from the region where the original image is located by matrix, and then they are spliced together to form a new sample image. The process of generating images and the new image samples are shown in Figure 15. The number of sample sets for Mosaic data enhancements is shown in Table 4.

After the above three data enhancement methods, the data set A was generated by the traditional data enhancement method, and the data set X was generated by the combination of the traditional data enhancement and the DCGAN data enhancement. The data set Y was constructed by the combination of the traditional data enhancement, the DCGAN and the Mosaic data enhancement, which was used as the three sample sets in the subsequent experiment. The sample situation of the data set is shown in Table 5.



**Figure 13.** Generates a picture rendering.

**Table 2.** Image data distribution of original data set and data set A.

Category	Sample number of original data set	Sample number of dataset A
tangerine peel images of 1 year	121	1089
tangerine peel images of 5 year	118	1062
tangerine peel images of 8 year	136	1224
tangerine peel images of 10 year	128	1152
tangerine peel images of 15 year	110	990
Aggregate	613	5517

**Table 3.** Number of DCGAN data enhancement samples.

Category	Sample number of original data set	Sample number of dataset B
tangerine peel images of 1 year	121	821
tangerine peel images of 5 year	118	818
tangerine peel images of 8 year	136	836
tangerine peel images of 10 year	128	828
tangerine peel images of 15 year	110	810
Aggregate	613	4113

Experimental design

Two groups of control experiments were conducted to verify the accuracy of the improved model in identifying the year of tangerine peel. The first group used the improved ResNet50 model to conduct a control experiment on the data sets A, X and Y. In addition to the original ResNet50 and TPR algorithm, the second group added VGG16 (Simonyan & Zisserman, 2014), InceptionV3 (Szegedy et al., 2016) and AlexNet (Krizhevsky et al., 2012) three models to conduct a control experiment on the data set Y. The data set was divided into training set, test set and verification set, and divided according to the ratio of 7:2:1. The batch-size was set to 16. The learning rate is 0.0002, and

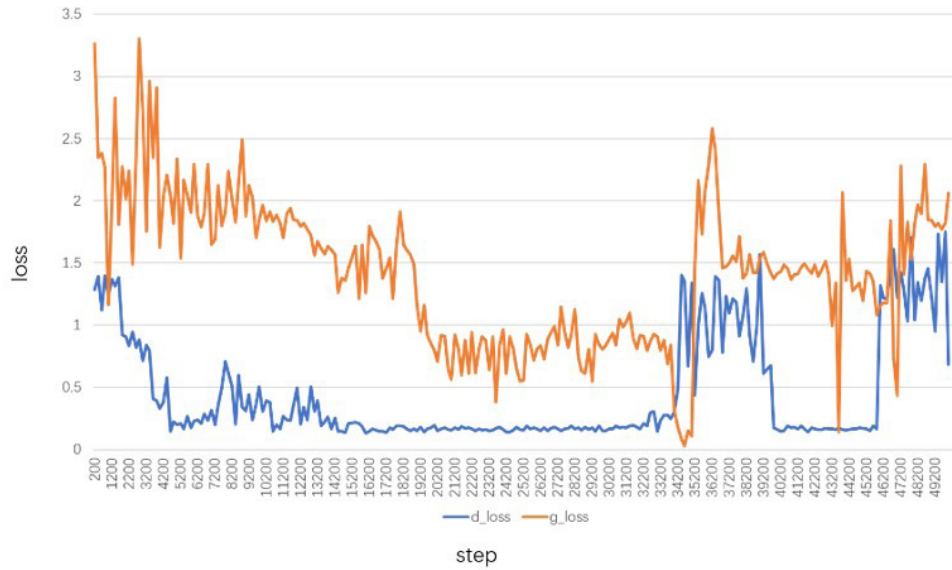


Figure 14. Generates a training loss curve.

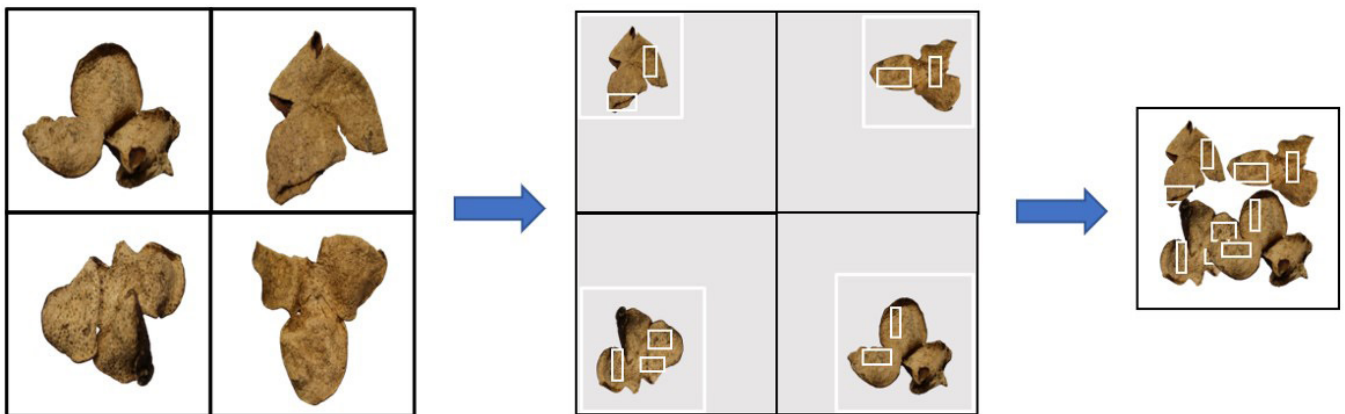


Figure 15. Mosaic generated images.

Table 4. Number of mosaic data enhanced samples.

Category	Sample number of original data set	Sample number of dataset C
tangerine peel images of 1 year	121	968
tangerine peel images of 5 year	118	944
tangerine peel images of 8 year	136	1088
tangerine peel images of 10 year	128	1024
tangerine peel images of 15 year	110	880
Aggregate	613	4904

Table 5. Data set image data distribution.

Category	Sample number of dataset A	Sample number of dataset X	Sample number of dataset Y
tangerine peel images of 1 year	1089	1910	2878
tangerine peel images of 5 year	1062	1880	2824
tangerine peel images of 8 year	1224	2060	3148
tangerine peel images of 10 year	1152	1980	3004
tangerine peel images of 15 year	990	1800	2680
Aggregate	5517	9630	14534

the sample input size is  $224 \times 224$ . All the models are iterative for 50 rounds.

#### 4.2 Analysis of effect

Through contrasting experiments, it was found that the data set Y performed best in the three data sets. Due to the

combination of three data enhancement modes, the data set Y had better sample richness and achieved better results in feature extraction. The accuracy rate was 98.8%, the accuracy rate was



98.7%, the recall rate was 98.9%, and the F1 value was 98.8%. Compared with the data sets A and X, the accuracy rate was increased by 1.5% and 0.6%, the accuracy rate was increased by 0.5% and 0.1%, the recall rate was increased by 1.1% and 0.4%, and the F1 value was increased by 0.9% and 0.3%, as shown in Table 6. The improved ResNet50 model reduced the amount of calculation and the training time. The training time was 46 s shorter than the original ResNet50, 99 s shorter than InceptionV3, and 177 s shorter than VGG16. The network depth was increased and the accuracy was improved by convolution kernel decomposition combined with CSPNet, the accuracy of the improved model was 98.8%, 1.1% higher than the original model, 0.4% higher than InceptionV3, 1.8% higher than VGG16 and 2.7% higher than AlexNet, as shown in Table 7. It was proved that the training speed and accuracy of the improved model were significantly improved, the overfitting problem was avoided, and the generalization ability of the model was enhanced.

Figure 16 shows the evaluation indexes of the identification effect of different models on different years of tangerine peel. The horizontal axis represents the simple name of the year of tangerine peel, and the vertical axis represents the value of the

**Table 6.** Comparative experimental results of three datasets.

Category	Accuracy	Precision	Recall	F1-score
Data set A	97.3%	98.2%	97.8%	97.9%
Data set X	98.2%	98.6%	98.5%	98.5%
Data set Y	98.8%	98.7%	98.9%	98.8%

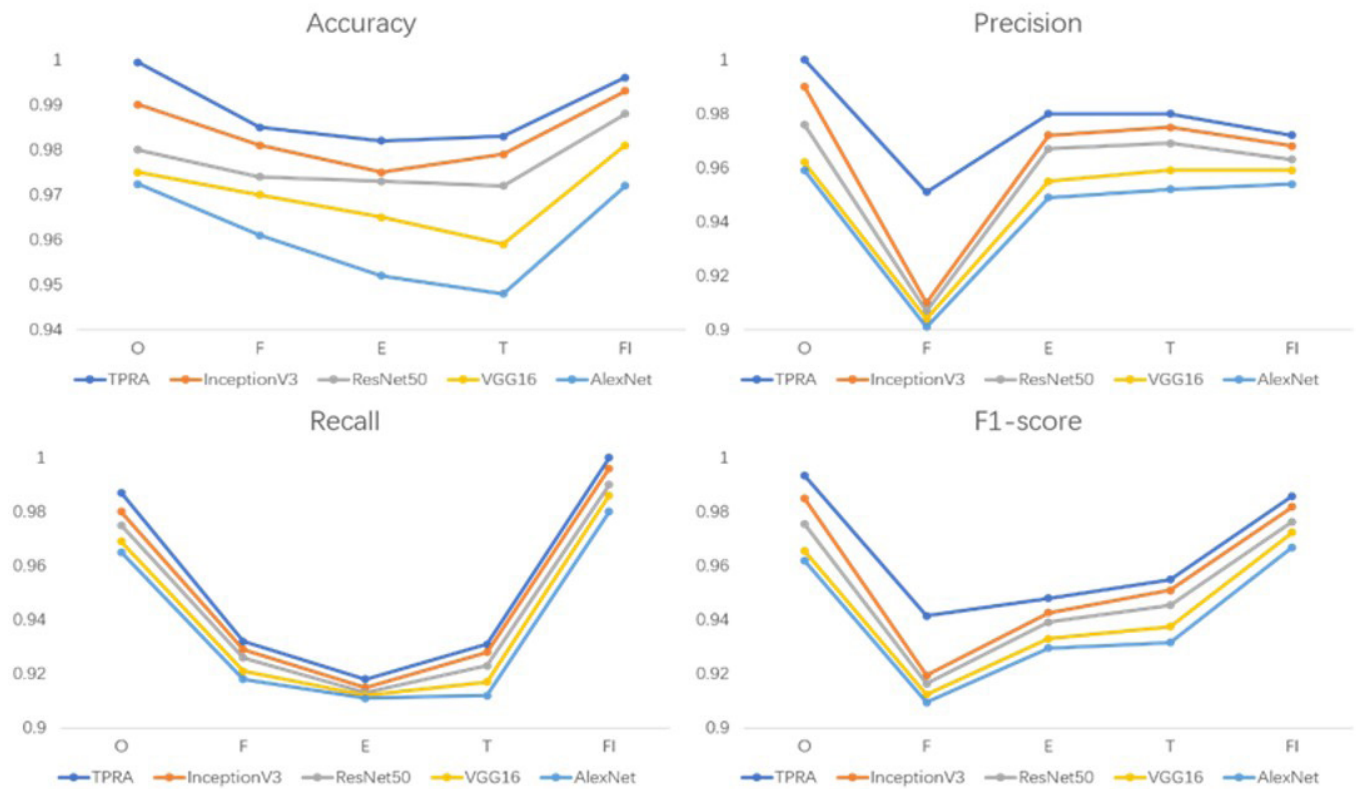
evaluation index. Considering the size, accuracy and training speed of the model parameters comprehensively, the effect of the improved ResNet50 model was the best. Moreover, the improved model was verified in data set Y. The accuracy, precision, recall and F1 values are shown in Table 8.

**Table 7.** Model comparison experiment results.

Model	Accuracy	Training time	Parameter size
TPRA	98.8%	455 s	96.3 MB
InceptionV3	98.4%	554 s	232.2 MB
ResNet50	97.7%	501 s	164.1 MB
VGG16	97.0%	632 s	642.6 MB
AlexNet	96.1%	309 s	94.4 MB

**Table 8.** Detailed classification results of peel years.

Category	Accuracy	Precision	Recall	F1-score
tangerine peel images of 1 year	99.4%	100.0%	98.7%	99.3%
tangerine peel images of 5 year	98.5%	95.1%	93.2%	94.1%
tangerine peel images of 8 year	98.2%	98.0%	91.8%	94.8%
tangerine peel images of 10 year	98.3%	98.0%	93.1%	95.5%
tangerine peel images of 15 year	99.6%	97.2%	100.0%	98.6%



**Figure 16.** Detailed classification results of different models.

## 5 Conclusion

Aiming at the problems of small number of samples, difficult feature extraction and difficult year identification, this paper proposes a new lightweight algorithm TPRA. TPRA builds the database through the hybrid data enhancement method, which combines the traditional data enhancement, DCGAN and Mosaic three data enhancement methods, and expands the number of samples in the original data set by about 23 times. The expanded data set shows superior performance in the identification of tangerine peel year. In the experiment, it is proved that this hybrid data enhancement method is superior to the single data enhancement method by comparing the accuracy, recall and F1 values. The performance of the model is improved and the parameter calculation is reduced. At the same time, TPRA introduces the attention mechanism module CBAM combined with the cross-stage local network CSPNet and adjusts the algorithm structure. The lightweight algorithm shortens the training time and improves the recognition accuracy. The average recognition accuracy was 98.8%, the average accuracy was 97.7%, the average recall rate was 95.4%, and the average F1 was 96.5%. It has important application value for the identification of the year of tangerine peel, and also has corresponding reference value for other Chinese medicine products.

## Acknowledgements

This work was supported by Shandong Provincial Natural Science Foundation (Project no.ZR2021ME018), Guangdong Province Key Field Research and Development Plan (Project no.2018B020241003-09), Shandong Province Agricultural Machinery Equipment Research and Development Innovation Plan (Project no.2018YZ002), National Peanut Industry Technology System (CARS-13-Mechanized Sowing and Field Management Positions), and National University Student Innovation and Entrepreneurship Training Plan (Project nos. S202010435013 and 201810435050X).

## References

- Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). *Yolov4: optimal speed and accuracy of object detection*. Retrieved from <https://arxiv.org/abs/2004.10934>
- Bottou, L. (2012). Stochastic gradient descent tricks. In G. Montavon, G. B. Orr & K.-R. Müller (Eds.), *Neural networks: tricks of the trade* (pp. 421-436). Berlin: Springer. [http://dx.doi.org/10.1007/978-3-642-35289-8\\_25](http://dx.doi.org/10.1007/978-3-642-35289-8_25).
- Chen, T.-C., & Yu, S.-Y. (2022). Research on food safety sampling inspection system based on deep learning. *Food Science and Technology*, 42, e29121. <http://dx.doi.org/10.1590/fst.29121>.
- Dileep, M. R., & Pournami, P. N. (2019). Ayurleaf: a deep learning approach for classification of medicinal plants. In K. Suresh (Ed.), *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)* (pp. 321-325). New York: IEEE. <http://dx.doi.org/10.1109/TENCON.2019.8929394>.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11), 139-144. <http://dx.doi.org/10.1145/3422622>.
- Gu, X.-h., Xu, R., Yuan, G.-l., Lu, H., Gu, B.-r., & Xie, H.-p. (2010). Preparation of chlorogenic acid surface-imprinted magnetic nanoparticles and their usage in separation of traditional Chinese medicine. *Analytica Chimica Acta*, 675(1), 64-70. <http://dx.doi.org/10.1016/j.aca.2010.06.033>. PMID:20708118.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In E. Mortensen & K. Saenko (Eds.), *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770-778). New York: IEEE.
- Ho, S.-C., & Kuo, C.-T. (2014). Hesperidin, nobiletin, and tangeretin are collectively responsible for the anti-neuroinflammatory capacity of tangerine peel (*Citri reticulatae* pericarpium). *Food and Chemical Toxicology*, 71, 176-182. <http://dx.doi.org/10.1016/j.fct.2014.06.014>. PMID:24955543.
- Hu, J.-L., Wang, Y.-K., Che, Z.-Y., Li, Q.-Q., Jiang, H.-K., & Liu, L.-J. (2020). Image recognition of Chinese herbal pieces based on multi-task learning model. In T. Ahn, S. Choi & P. Veltri (Eds.), *2020 IEEE International Conference on Bioinformatics and Biomedicine* (pp. 1555-1559). New York: IEEE. <http://dx.doi.org/10.1109/BIBM49941.2020.9313412>.
- Janani, R., & Gopal, A. (2013). Identification of selected medicinal plant leaves using image features and ANN. In P. B. Prasad, R. Verma & C. Shekhar (Orgs.), *2013 International Conference on Advanced Electronic Systems* (pp. 238-242). New York: IEEE. <http://dx.doi.org/10.1109/ICAES.2013.6659400>.
- Kan, H. X., Jin, L., & Zhou, F. L. (2017). Classification of medicinal plant leaf image based on multi-feature extraction. *Pattern Recognition and Image Analysis*, 27(3), 581-587. <http://dx.doi.org/10.1134/S105466181703018X>.
- Kingma, D. P., & Ba, J. (2014). *Adam: a method for stochastic optimization*. Retrieved from <https://arxiv.org/abs/1412.6980>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097-1105.
- Lin, L., Liu, Z. X., & Mo, Y. Y. (2008). Dynamic analysis of the total flavone and the hesperidin from different specific years in XinHui dried tangerine peel. *Lishizhen Medicine and Materia Medica Research*, 19(6), 1432-1433.
- Liu, S., Chen, W., & Dong, X. (2018). Automatic classification of Chinese herbal based on deep learning method. In A. Roy, G. Xiao & L. Rutkowski (Eds.), *2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery* (pp. 235-238). New York: IEEE. <http://dx.doi.org/10.1109/FSKD.2018.8687165>.
- Mao, R., He, J., Shao, Z., Yarlagadda, S. K., & Zhu, F. (2021). Visual aware hierarchy based food recognition. In A. Bimbo, R. Cucchiara, S. Sclaroff, G. M. Farinella, T. Mei, M. Bertini, H. J. Escalante, & R. Vezzani (Eds.), *Pattern recognition. ICPR International Workshops and Challenges: virtual event, January 10–15, 2021, proceedings, part V* (pp. 571-598). Cham: Springer. [http://dx.doi.org/10.1007/978-3-030-68821-9\\_47](http://dx.doi.org/10.1007/978-3-030-68821-9_47).
- Mish, M. D. (2019). *Mish: a self regularized non-monotonic neural activation function*. Retrieved from <https://arxiv.org/abs/1908.08681v1>
- Mukti, I. Z., & Biswas, D. (2019). Transfer learning based plant diseases detection using ResNet50. In *2019 4th International Conference on Electrical Information and Communication Technology* (pp. 1-6). New York: IEEE. <http://dx.doi.org/10.1109/EICT48899.2019.9068805>.
- Radford, A., Metz, L., & Chintala, S. (2015). *Unsupervised representation learning with deep convolutional generative adversarial networks*. Retrieved from <https://arxiv.org/abs/1511.06434>
- Shijie, J., Ping, W., Peiyi, J., & Siping, H. (2017). Research on data augmentation for image classification based on convolution neural

- networks. In *2017 Chinese Automation Congress* (pp. 4165-4170). New York: IEEE. <http://dx.doi.org/10.1109/CAC.2017.8243510>.
- Simonyan, K., & Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition*. Retrieved from <https://arxiv.org/abs/1409.1556>
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In E. Mortensen & K. Saenko (Eds.), *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2818-2826). New York: IEEE. <http://dx.doi.org/10.1109/CVPR.2016.308>.
- Wang, C.-Y., Liao, H.-Y. M., Wu, Y.-H., Chen, P.-Y., Hsieh, J.-W., & Yeh, I.-H. (2020). CSPNet: a new backbone that can enhance learning capability of CNN. In E. Mortensen & M. Masson (Eds.), *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 1571-1580). New York: IEEE. <http://dx.doi.org/10.1109/CVPRW50498.2020.00203>.
- Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). Cbam: convolutional block attention module. In V. Ferrari, M. Hebert, C. Sminchisescu & Y. Weiss (Eds.), *Computer vision – ECCV 2018: 15th European conference, Munich, Germany, September 8–14, 2018, proceedings, part VII* (pp. 3-19). Cham: Springer. [http://dx.doi.org/10.1007/978-3-030-01234-2\\_1](http://dx.doi.org/10.1007/978-3-030-01234-2_1).
- Xing, C., Huo, Y., Huang, X., Lu, C., Liang, Y., & Wang, A. (2020). Research on image recognition technology of traditional Chinese medicine based on deep transfer learning. In T. Yang (Org.), *2020 International Conference on Artificial Intelligence and Electromechanical Automation* (pp. 140-146). New York: IEEE. <http://dx.doi.org/10.1109/AIEA51086.2020.00037>.
- Xu, Y., Wen, G., Hu, Y., Luo, M., Dai, D., Zhuang, Y., & Hall, W. (2021). Multiple attentional pyramid networks for Chinese herbal recognition. *Pattern Recognition*, 110, 107558. <http://dx.doi.org/10.1016/j.patcog.2020.107558>.
- Yi, L., Dong, N., Liu, S., Yi, Z., & Zhang, Y. (2015). Chemical features of pericarpium Citri reticulatae and pericarpium Citri reticulatae Viride revealed by GC-MS metabolomics analysis. *Food Chemistry*, 186, 192-199. <http://dx.doi.org/10.1016/j.foodchem.2014.07.067>. PMID:25976810.
- Zhang, P., & Xu, F. (2022). Effect of AI deep learning techniques on possible complications and clinical nursing quality of patients with coronary heart disease. *Food Science and Technology*, 42, e42020. <http://dx.doi.org/10.1590/fst.42020>.
- Zou, Z., Wang, L., Chen, J., Long, T., Wu, Q., & Zhou, M. (2022). Research on peanut variety classification based on hyperspectral image. *Food Science and Technology*, 42, e18522. <http://dx.doi.org/10.1590/fst.18522>.