



Jujube defect recognition method based on boosted convolutional neural network

Chenchen DONG¹ , Mao PANG^{1*}, Miaolong CAO¹

Abstract

In order to solve the problem of difficult and slow identification of jujube defects, a convolutional neural network model based on boosted EfficientNetv2 was proposed by taking the dry, cracked, broken and normal jujube in jujube as the research object. First, optimize the model structure, and set of the first Fused_MBCConv in EfficientNetv2 3*3 convolution improved to parallel 1*1 convolution, 3*3 convolution and two serial 3*3 convolution. With reference to the CSPNet idea, one part of the convolved feature map in the MBCConv module is directly spliced across channels, the other part is output through the dense block through the transition layer, and then spliced with the feature map in the first part; Then, the original Swish activation function is replaced by the optimal FReLU activation function; Finally, the Coordinate Attention module is introduced to embed the position information into the channel attention to optimize the model. The experimental results showed that the recognition rates of dry jujube, cracked jujube, broken jujube and normal jujube were 95.32%, 98.79%, 98.19% and 97.81% respectively, and the average recognition rate was 97.39%. Compared with other algorithms, the model has faster speed and higher recognition accuracy for defective jujube.

Keywords: jujube; defect identification; convolutional neural network; activation function; attention mechanism.

Practical Application: Application of convolutional neural network in defect detection of jujube.

1 Introduction

Jujube has a long history of planting in China and has a wide variety. It has been regarded as a good tonic since ancient times. It has the effects of tonifying the spleen and stomach, supplementing qi and blood, calming the nerves, and easing the drug properties. It is also highly commercialized. With the continuous upgrading of people's consumption capacity and the gradual popularization of food health knowledge, the demand for red dates as nutritional health food has grown rapidly, driving the rapid development of jujube related processing product industry. The jujube food processing industry includes preserved fruits, beverage granules, cans, jams, cakes, health food, enzymes, freeze-dried food and other kinds of food. In the future, with the continuous development of jujube deep processing technology, China's jujube market will meet more consumer demand, and the market consumption will continue to expand. The sorting of jujube quality can enhance the added value of products in the market, and is also an important factor affecting the price and sales of jujube. After appearance quality grading, the value of high-quality jujube can be boosted, while defective jujube can be processed and transformed into feed, edible yeast and other products.

During the growth, picking and transportation of jujube, it is inevitable that there will be defects such as dry strips, rags and cracks. These defects lead to uneven quality of jujube. Therefore, it is essential to accurately identify and screen defective jujube before processing. At present, the surface defects of jujube are mainly observed manually before they enter the market. This way is labor-intensive and subject to subjective influence, resulting in uneven sorting. The uneven quality of jujube will

directly affect the subsequent processing of jujube. The essence of convolutional neural network is multi-layer perceptron, which maps the input nonlinearity to the output. It uses local connection and weight sharing to reduce the number of weights and reduce the complexity of the model. Compared with the traditional feature extraction methods, the convolution layer contained in the convolution neural network has more powerful feature learning and abstract representation capabilities, and has automatic learning features, which is particularly suitable for the application of computer vision in food science. Therefore, convolutional neural network is widely used in food science research. Wen et al. (2020), proposed a classification algorithm based on convolution neural network to recognize the surface defects and textures of red dates. The feature map obtained after preprocessing the G component map of the red date RGB color map is used as the input of the network. The learning depth of the neural network is expanded by residual learning, and the activation function of the convolution neural network, ReLU, is replaced by SELU, The loss function softmax loss is replaced by center loss. The Dropout layer is introduced during training to reduce the risk of network over fitting, and the phenomenon of gradient dispersion and explosion in the network is solved as the learning depth deepens. The results show that the accuracy of the classification method is 96.11%. Jiang et al. (2019), studied and analyzed soybean leaf spot, mosaic, downy mildew and gray spot with the common soybean disease pictures as samples. The training set of the neural network model was obtained by preprocessing the disease pictures such as binarization and contour segmentation. On this basis, the model was optimized

Received 29 Nov., 2022

Accepted 19 Jan., 2023

¹ College of Mechanical and Energy Engineering, Zhejiang University of Science and Technology, Hangzhou, Zhejiang Province, China

*Corresponding author: palmy@zust.edu.cn

in many aspects, and a disease detection system for soybean was designed using convolution neural network technology. Jiang et al. (2022), proposed a cucumber disease identification method based on convolutional neural network. By collecting the sample pictures with disease characteristics, carrying out image enhancement processing, the cucumber leaf disease data set was made, and the disease identification effects of three different depth network models, namely AlexNet, VGG-16 and ResNet50, were studied. By designing different schemes for model training, the network model with the best training effect was found and the disease pictures were detected. The system can meet the expected requirements of cucumber disease identification, and has a high recognition accuracy. Based on the “Bagging” integrated learning method, Zeng et al. (2019) used E-CNN to build a basic convolutional neural network tree model through training sets. Then, according to the output results of each basic tree model, he obtained the target location method using the shortest path search between frames through the “voting” method, making the model location accuracy 100% and the recognition accuracy 98.48%. Although these methods have achieved good results in jujube defect recognition, traditional convolutional neural network and image processing have a large number of redundant parameters for jujube defect recognition, resulting in the recognition of defective jujube is not fast enough. At the same time, the complexity and diversity of jujube surface texture features and the uneven lighting during shooting will adversely affect the extraction of features based on color, texture, etc. In the actual industry, a more efficient and accurate algorithm model is needed to improve the speed of jujube defect recognition. To solve this problem, this paper proposes a convolutional neural network based on boosted EfficientNetv2 (Tan & Le, 2021), which reduces the number of model parameters and improves the detection accuracy of the model.

2 Materials and methods

2.1 Classification criteria

According to the grading standard of Chinese dried jujube (GB/T5835 of National Standards of the Republic of China, 2009), the defects of dried jujube refer to the fruits damaged by pests, machinery and chemicals during the growth, development and picking process. In this paper, the typical cracked, dry and broken jujube are taken as the objects to study the identification methods of defective jujube. Among them, the dry bar grade

refers to the dried immature fresh jujube, the fruit is hard and thin, the flesh is not full, the texture is hard, the peel color is light to yellow, and there is no luster, as shown in Figure 1a; Broken grade refers to the fruit damaged by pests, which damages the pulp, or leaves insect flocs, insect bodies and excreta around the fruit core, as shown in Figure 1b; The crack level refers to the cracking of jujube peel caused by natural cracking or mechanical damage during its growth, as shown in Figure 1c. Therefore, it is divided into four types of jujube data sets, namely, dried jujube, crackle jujube, broken jujube and normal jujube.

2.2 EfficientNetv2

EfficientNet combines and balances the three key dimensions of depth, width and resolution, uniformly scales the three dimensions through a set of fixed scaling coefficients, and explores the impact of the three coefficients on model performance (Tan & Le, 2019). EfficientNetv2 proposes an boosted progressive learning method, which can dynamically adjust the regularization method according to the size of the training image, and is superior to EfficientNet in training speed and number of parameters. EfficientNetv2 is mainly composed of mobile flip bottleneck convolutional structure (MBConv) and fused mobile flip bottleneck convolutional structure (Fused_MBConv) modules (Gupta & Akin, 2020). Its structure is shown in Figure 2.

MBConv consists of two 1×1 convolution, SE attention module, 3×3 Depth separable convolution (DW Conv) and residual edge composition. The depth separable convolution greatly reduces the parameters of the model. Fused_MBConv uses 3×3 convolution replaces 3×3 depth separable convolution and 1×1 convolution in MBConv. Compared with MBConv, Fused_MBConv has more parameters, but it has stronger feature extraction ability and graph reasoning speed. The best combination of Fused_MBConv and MBConv is obtained through the neural architecture search (NAS) method, so as to give full play to the advantages of the two modules and achieve the balance of precision, model parameters and reasoning speed. The structure parameters of EfficientNetv2 under this combination are shown in Table 1.

2.3 Improvement based on EfficientNetv2

In the EfficientNetv2 network model, the first Fused_MBConv has no extended convolution in the module. Although the serial convolutional structure can reduce memory access overhead,

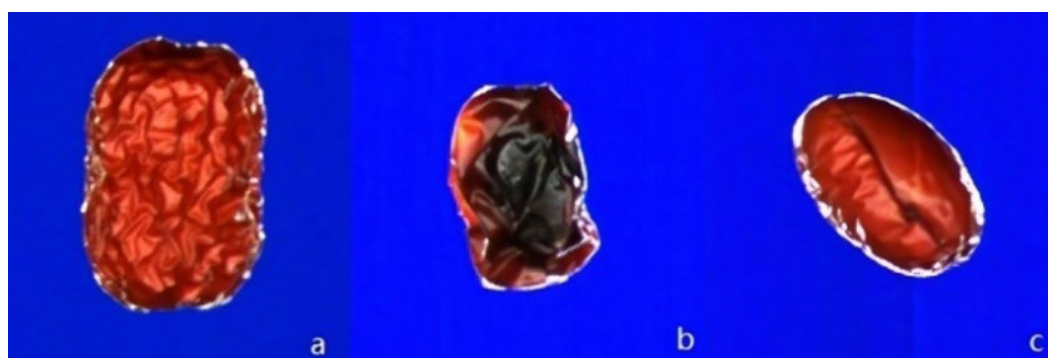


Figure 1. Example of jujube defect image (a: dried jujube, b: broken jujube, c: crackle jujube).

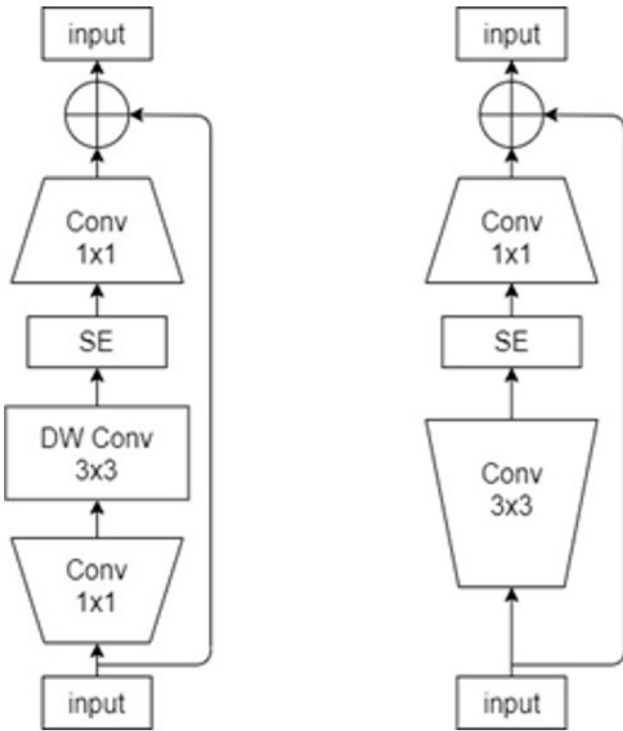


Figure 2. MBConv and Fused_MBConv structure diagram.

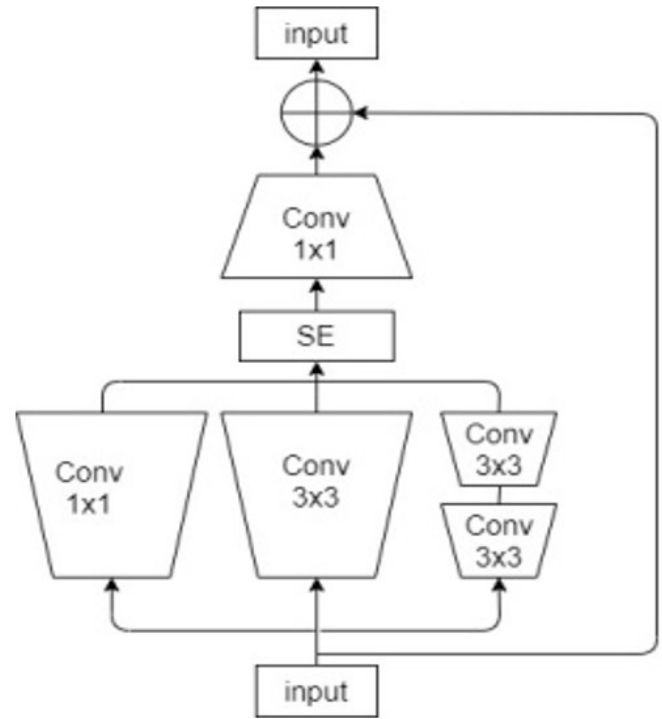


Figure 3. Boosted Fused_MBConv structure diagram.

Table 1. Efficient Netv2 network structure parameters.

Stage	Operator	Stride	Channels	Layers
1	Conv3*3	2	24	1
2	Fused-MBConv1,k3*3	1	24	2
3	Fused-MBConv4,k3*3	2	48	4
4	Fused-MBConv4,k3*3	2	64	4
5	MBConv4,k3*3,SE0.25	2	128	6
6	MBConv6,k3*3,SE0.25	1	160	9
7	MBConv6,k3*3,SE0.25	2	272	15
6	Conv2D&Pooling&FC	-	1792	1

for some subtle features of red dates, only 3*3 convolution, the extracted features are insufficient. In order to further enrich the model and extract more image features, the idea of multi-scale convolution is used to convert the 3*3 convolution boosted to parallel 1*1 convolution, 3*3 convolution and 5*5 convolution, which can enrich the feature extraction of objects of different sizes. At the same time, 5*5. The parameter quantity of convolutional kernel is too large. In order to further reduce the parameter quantity and improve the operation speed 5*5 convolution is changed to two serial 3*3 convolution. After multi-scale convolution, the feature map is spliced on the channel dimension, and the subsequent SE attention is fused and 1*1 convolution cross channel information exchange and fusion. The improved Fused_MBConv module structure is shown in Figure 3.

The latter stage of EfficientNetv2 model adopts the MBConv structure. Because the idea of multi-scale convolution is used in the Fused_MBConv structure in the former stage of the model, the original single path convolution is changed to three path convolution, which greatly increases the model parameters

and computation. In order to further improve the applicability of the model, it is necessary to reduce the amount of model parameters and calculations. In the MBConv module, using the idea of the Cross Stage Partial Network(CSPNet) (Wang et al., 2020). One part of the convolved feature map in the MBConv module is directly spliced across channels, and the other part is output by the Dense Block through the transition layer, and then spliced with the feature map in the first part. With this step, the convolved parameters in the model can be reduced, The gradient changes are integrated into the feature map from beginning to end, which enriches the gradient combination of the model and enhances the learning ability of the network. The improved MBConv module structure is shown in Figure 4.

Attention module through parameter adjustment, the attention module strengthens the network's attention to the important features in the data and suppresses the background features in the data to improve the segmentation accuracy of the model recognition results, especially for the improvement of details. In the jujube defect recognition image, some defects are often not obvious. Not only the shadows of wrinkles and dark parts should be considered, but also the segmentation of fine parts such as cracks should be considered. Therefore, the introduction of attention mechanism can help to refine the segmentation and improve the recognition accuracy. Common attention mechanisms are SENet and ECANet. And Coordinate Attention can not only capture cross channel information, but also capture directional awareness and location awareness information, help the model more accurately locate and identify the target of interest, and enhance features by strengthening information representation (Hou et al., 2021). The structure of Coordinate Attention is shown in Figure 5.

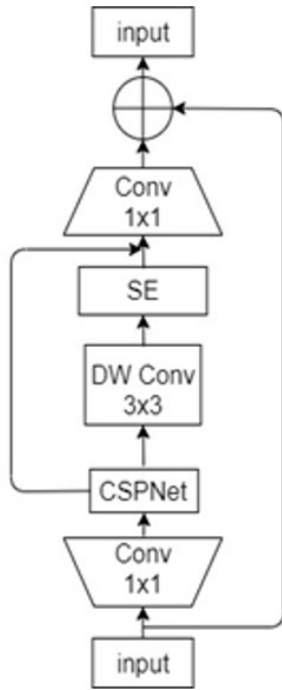


Figure 4. Boosted MBConv structure diagram.

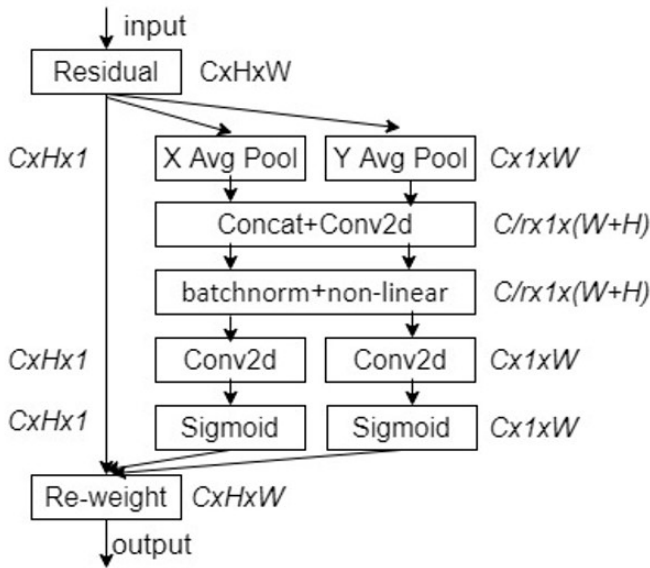


Figure 5. Coordinate Attention structure diagram.

The coordinate attention module performs average pooling in the horizontal and vertical directions respectively to obtain two one-dimensional vectors, and then performs splicing operations in the spatial dimension and passes 1*1 convolution is used to compress the channel, and then batch normalization and nonlinear relationship are used to encode the spatial information in the vertical and horizontal directions. After separation, they pass 1*1 convolution to obtain the same channel number as the input feature map, and finally normalized weighting.

Coordinate Attention encodes channel relationship and long-term dependence through accurate position information,

which is divided into two steps: information embedding and generation. The global pooling method is usually used for the global encoding of the channel attention encoding spatial information, but it is difficult to save the location information because it compresses the global spatial information into the channel descriptor. In order to enable the attention module to capture the remote spatial interaction with accurate location information, the global pooling is transformed into a pair of one-dimensional feature codes.

For a given input X, each channel is coded along the horizontal and vertical coordinate directions using the pooled kernel of dimensions (H, 1) and (1, W). The output of the cth channel with height h is (Equation 1):

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} X_c(h, i) \quad (1)$$

The output of the cth channel with width w is (Equation 2):

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} X_c(j, w) \quad (2)$$

The above two transformations aggregate features along two spatial directions respectively to obtain a pair of directional perception feature maps, and allow the attention module to capture the long-term dependency relationship along one spatial direction and save the accurate location information along the other spatial direction, which helps the network locate the target of interest more accurately.

Coordinate Attention generation is to make better use of the above transformation to obtain the global receptive field, and code the representation generated by accurate position information. Its operation starts with cascading two feature maps generated by modules, and then uses a shared 1*1 convolution is transformed to generate $f \in R^{C/r \times (H+W)}$ is the intermediate feature map of spatial information in the horizontal and vertical directions. r represents the down sampling ratio, which is used to control the size of the module as the SE module. Next, divide f into two independent tensors $f^h \in R^{C/r \times H}$ and $f^w \in R^{C/r \times W}$ along the spatial dimension, reuse two 1*1 convolution F_h and F_w Transform the characteristic graphs f^h and f^w to the same number of channels as input X (Zhou et al., 2022). The formula is (Equation 3):

$$g^h = \sigma(F_h(f^h)), g^w = \sigma(F_w(f^w)) \quad (3)$$

Where: σ is a sigmoid function. In order to reduce the complexity of the network, an appropriate down sampling ratio r is used to reduce the number of channels, and then g^h and g^w are expanded as attention weights. Finally, the output of the Coordinate Attention block is expressed as follows (Equation 4):

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (4)$$

The Coordinate Attention module efficiently integrates the spatial coordinate information into the generated feature map by embedding the position information into the channel attention

and decomposing the channel attention into two parallel one-dimensional feature codes. Each feature map captures the remote dependencies of the input feature map along a spatial direction.

In order to achieve pixel level spatial information modeling capability in the activation function stage and improve accuracy, this paper replaces the Swish activation function in the original model with the FReLU activation function that performs well in defect recognition and image classification tasks (Ramachandran et al., 2017; Ma et al., 2020). The calculation is (Equation 5):

$$FReLU(x) = \max(x, T(x)) \quad (5)$$

Where: $T(x)$ is a two-dimensional spatial condition. Using the characteristics of spatial conditions $T(x)$ and max functions, FReLU can provide pixel level modeling capabilities and spatial layout capabilities, naturally extract the spatial structure of objects, and generate spatial dependence while nonlinear activation of FReLU, solving the spatial insensitivity of other activation functions. The schematic diagram of the FReLU activation function is shown in Figure 6.

The two-dimensional space condition $T(x)$ in Figure 6 is implemented by a depthWise separable conv and batchnorm. The input channel and output channel of the depth separable conv are the same.

3 Results and discussion

3.1 Ablation test

The attention mechanism and activation function ablation test were trained on the jujube image training data set. Based on the boosted EfficientNetv2 network, the attention mechanism SENet module, ECANet module, Coordinate Attention module, Swish activation function and FReLU activation function were separately introduced, while the other parts were consistent to explore the impact on the experimental effect. Table 2 shows the ablation test of each network model and the original model introducing attention mechanism and activation function on the training dataset.

It can be seen from Table 3 that after the attention mechanism is increased, the impact of redundant information in the image is reduced through selective attention to features. In particular, Coordinate Attention can capture cross channel information, direction awareness and location awareness information, and can maximize the use of effective feature information. Compared with the introduction of the SENet module, the introduction of the Coordinate Attention module has an increase of 0.56% in mAP, 0.20% in ECANet module and 0.26% in EfficientNetv2 model.

Because the details of the defective part and the normal part in the jujube surface image are not very different, it

will interfere with the training of the convolutional neural network. When the boosted EfficientNetv2 is introduced into the ECA module, the accuracy rate decreases by 1.86%, which indicates that the attention module can not effectively guide the network to focus on the defect differentiation area, ignoring the differences between different categories. The introduction of the Coordinate Attention module can take more consideration of the distinguishable areas of details, and guide the network to optimize the correct weight, Greatly promote the distinguishable local features of online learning for detailed classification, and effectively avoid the negative impact of ECA attention module. Therefore, choosing Coordinate Attention mechanism can further enhance the network's feature extraction ability and greatly improve the network performance.

After the activation function is added, the mAP value of the boosted EfficientNetv2 network tested on the training data set is boosted. The FReLU activation function achieves the pixel level spatial information modeling ability, and performs the best in the jujube defect detection task. After the introduction

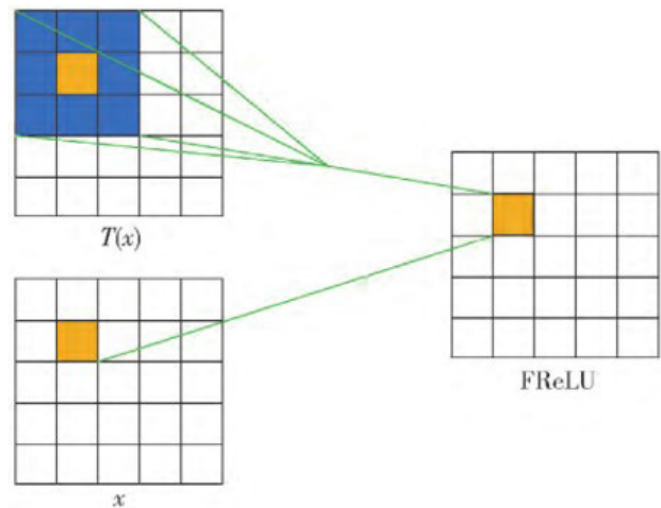


Figure 6. Schematic diagram of FReLU activation function.

Table 2. Comparison of results of using different attention mechanisms and activation functions in efficientNetv2 network.

Attention	Activation function	model	mAP (%)
-	-	EfficientNetv2	95.29
SENet	-	EfficientNetv2	94.99
ECANet	-	EfficientNetv2	95.35
Coordinate Attention	-	EfficientNetv2	95.55
-	Swish	EfficientNetv2	95.64
-	FReLU	EfficientNetv2	96.47

mAP: Mean Average Precision.

Table 3. Comparison of different categories before and after improvement of EfficientNetv2 model.

Model	Dried jujube accuracy(%)	Crackle jujube accuracy(%)	Broken jujube accuracy(%)	Normal jujube accuracy(%)	mAP (%)	FPS
EfficientNetv2	93.21	97.25	96.18	95.17	95.29	25.3
Boosted EfficientNetv2	95.32	98.79	98.19	97.81	97.39	24.1

mAP: Mean Average Precision; FPS: Frames Per Second.

of FReLU activation function, compared with the introduction of Swish activation function on the mAP, it has increased by 0.83%. The FReLU activation function provides pixel level modeling capability and spatial layout capability, and naturally extracts the spatial layout of objects, which greatly improves the network performance. Therefore, this model selects FReLU activation function.

3.2 Result analysis

EfficientNetv2 model and boosted EfficientNetv2 model (improving model structure, introducing Coordinate Attention module and replacing FReLU activation function) were trained on the jujube validation data set respectively. The accuracy and average accuracy of each category are shown in Table 3.

The boosted Efficient Netv2 model has boosted the recognition accuracy of all categories to a certain extent, with an average accuracy rate of 97.39%, which is 2.09% higher than the original Efficient Netv2 model. This shows that it is effective to optimize the model by optimizing the model structure with rich features, using the FReLU activation function and embedding the location information into the channel attention. In terms of model detection rate, the boosted model transmits 24.1 frames per second, which is somewhat faster than the original model. In general, the boosted model has significantly boosted the detection speed, which can save time and cost, and has strong real-time detection.

3.3 Model validation

In order to verify the advantages of the boosted EfficientNetv2 model in jujube defect recognition, the proposed algorithm is compared with the classic image classification networks of AlexNet, VGG16, ResNet101, Inception v3, MobileNetv2, EfficientNetv2 (Russakovsky et al., 2015; Simonyan & Zisserman, 2014; He et al., 2016; Szegedy et al., 2015; Sandler et al., 2018). Use it to identify 976 pictures in the test set, and the comparison results of mAP and FPS of each category for each model are shown in Table 4.

It can be seen from Table 4 that the average recognition accuracy of EfficientNetv2 model on the test set is 98.37%, which is significantly boosted compared with the original model. Compared with other deep learning image classification networks, EfficientNetv2 model has a higher recognition accuracy and a slight advantage in recognition speed. Through experiments, it is verified that the boosted EfficientNetv2 model has good generalization.

Table 4. Test results of seven different models on jujube test data set.

Model	mAP(%)	FPS
AlexNet	80.02	34.2
VGG16	83.94	92.3
ResNet101	93.13	85.4
Inceptionv3	93.97	43.2
MobileNetv2	91.15	37.1
EfficientNetv2	94.59	40.3
boosted EfficientNetv2	98.37	36.7

mAP: Mean Average Precision; FPS: Frames Per Second.

4 Conclusion

Convolutional neural network can effectively overcome the disadvantage that the detection results are easily affected by samples and human subjectivity. It has a good adaptability to the spatial invariability of the image. Its powerful self-learning ability can automatically extract and learn the feature information in the image without additional supervision and training. In recent years, it has been gradually applied to defect detection, image recognition and other fields (Abdelbasset et al., 2022; Zhu et al., 2022). In traditional machine learning, the sample features are mainly obtained by means of feature transformation. For different tasks, extracting features of different modes is not universal. And traditional machine learning can only use relatively simple functional form. The expression ability of models directly affects the final prediction effect of machine learning models, while simple functions usually do not have the expression ability of complex models. If we simply expand the function family and use more complex functions to learn, then the model is often prone to over fitting and the generalization ability is reduced. Compared with the traditional machine learning method, the convolutional neural network method can save reasoning and create new reasoning on subsequent layers through layer transfer learning; Before using convolutional neural network algorithm, feature extraction is not necessary, but is completed in the training process; It has stronger ability of feature learning and abstract representation.

In this experiment, EfficientNetv2 network is used as the backbone network of jujube defect recognition feature extraction. By optimizing the model structure, using better FReLU activation function and introducing Coordinate Attention mechanism, the model can accurately and quickly identify the typical defective jujube of dry, cracked and rotten jujube.

Through the test, the comprehensive recognition accuracy of jujube has increased by 3.78%, reaching 98.37%. Compared with other image classification networks, the boosted EfficientNetv2 network model performs better in defect recognition, with higher classification accuracy and faster recognition rate. The model of this experiment can be embedded into the machine vision jujube sorting equipment to replace manual identification, and realize the defect screening of different jujube external quality, which not only saves labor, but also is fast and accurate, and is not affected by human factors. This model also has certain application value for the surface defect detection of other fruits and crops, and is helpful for the intelligent and industrialized development of agricultural products.

The boosted EfficientNetv2 model can realize the detection and sorting of jujube surface defects, but the internal parameters such as moisture content, hardness and sedimentation value cannot be detected. Near infrared hyperspectral nondestructive testing and classification technology can be used to detect the internal quality of jujube, which can be used as the next stage of this experiment (Zhou et al., 2023).

Acknowledgements

The research was funded by “the Natural Science Foundation of Zhejiang Province, China” (LQY19E050001).

References

- Abdelbasset, W. K., Nambi, G., Elkhohli, S. M., Eid, M. M., Alrawaili, S. M., & Mahmoud, M. Z. (2022). Application of neural networks in predicting the qualitative characteristics of fruits. *Food Science and Technology (Campinas)*, 42, e118821. <http://dx.doi.org/10.1590/fst.118821>.
- Gupta, S., & Akin, B. (2020). Accelerator-aware neural network design using automl. *arXiv*, arXiv:2003.02838.
- He, K., Zhang, X., & Ren, S. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. USA: IEEE.
- Hou, Q., Zhou, D., & Feng, J. (2021). Coordinate attention for efficient mobile network design. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. USA: IEEE. <http://dx.doi.org/10.1109/CVPR46437.2021.01350>.
- Jiang, F. Q., Li, Y., Yu, D. W., Sun, M., & Zhang, E. B. (2019). Soybean disease detection system based on Caffe convolution neural network. *Zhejiang Journal of Agriculture*, 31(7), 1177-1183.
- Jiang, L. S., Dong, Z. X., Hu, X., & Liu, Z. Q. (2022). Research on cucumber disease identification based on convolutional neural network. *Jisuan Jishu Yu Zidonghua*, 41(2), 153-157.
- Ma, N., Zhang, X., & Sun, J. (2020). Funnel activation for visual recognition. In *European Conference on Computer Vision*. GER: Springer.
- National Standards of the Republic of China. (2009). *GB/T 5835-2009: dried Chinese jujubes*. China: CNS Standards.
- Ramachandran, P., Zoph, B., & Le, Q. V. (2017). Searching for activation functions. *arXiv*, arXiv:1710.05941.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3), 211-252. <http://dx.doi.org/10.1007/s11263-015-0816-y>.
- Sandler, M., Howard, A., & Zhu, M. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. USA: IEEE.
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv*, arXiv:1409.1556.
- Szegedy, C., Liu, W., & Jia, Y. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. USA: IEEE.
- Tan, M., & Le, Q. (2019). Efficientnet: rethinking model scaling for convolutional neural networks. In *International conference on machine learning*. New USA: PMLR.
- Tan, M., & Le, Q. (2021). Efficientnetv2: smaller models and faster training. In *International Conference on Machine Learning*. USA: PMLR.
- 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. USA: IEEE. <http://dx.doi.org/10.1109/CVPRW50498.2020.00203>.
- Wen, H. X., Wang, J. J., & Han, F. (2020). Research on defect detection and classification method of jujube based on improved residual network. *Food and Machinery*, 36(1), 161-165.
- Zeng, T., Wu, J., & Ma, B. X. (2019). Jujube location and defect detection based on inter frame path search and E-CNN. *Mashin/Ha-Yi Kishavarzi*, 50(2), 307-314.
- Zhou, C. H., Xia, X. D., & Zhou, D. D. (2022). Pedestrian recognition by fusing grid mask and residual coordinate attention. *Microelectronics and Computer*, 39(05), 30-38.
- Zhou, M., Long, T., Zhao, Z. Y., Chen, J., Wu, Q. S., Wang, Y., & Zou, Z. Y. (2023). Honey quality detection based on near-infrared spectroscopy. *Food Science and Technology (Campinas)*, 43, e98822. <http://dx.doi.org/10.1590/fst.98822>.
- Zhu, Y. L., Ma, Z., Han, M., Li, Y., Xing, L., Lu, E., & Gao, H. (2022). Quantitative damage detection of direct maize kernel harvest based on image processing and BP neural network. *Food Science and Technology (Campinas)*, 42, e54322. <http://dx.doi.org/10.1590/fst.54322>.