



OPEN ACCESS

EDITED BY

Isabelle Piot-Lepetit,
INRAE Occitanie Montpellier, France

REVIEWED BY

Seyed-Hassan Miraei Ashtiani,
Dalhousie University, Canada
Aibin Chen,
Central South University Forestry and
Technology, China

*CORRESPONDENCE

Jing He
✉ hejing642023@126.com

RECEIVED 09 October 2023

ACCEPTED 09 January 2024

PUBLISHED 24 January 2024

CITATION

Wen J and He J (2024) Agricultural
development driven by the digital economy:
improved EfficientNet vegetable quality
grading.
Front. Sustain. Food Syst. 8:1310042.
doi: 10.3389/fsufs.2024.1310042

COPYRIGHT

© 2024 Wen and He. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Agricultural development driven by the digital economy: improved EfficientNet vegetable quality grading

Jun Wen and Jing He*

School of Agriculture, Guangxi University, Nanning, China

Introduction: The conventional manual grading of vegetables poses challenges that necessitate innovative solutions. In this context, our paper proposes a deep learning methodology for vegetable quality grading.

Methods: To address the scarcity of vegetable datasets, we constructed a unique dataset comprising 3,600 images of diverse vegetables, including lettuce, broccoli, tomatoes, garlic, bitter melon, and Chinese cabbage. We present an improved CA-EfficientNet-CBAM model for vegetable quality grading. The CA module replaces the squeeze-and-excitation (SE) module in the MobileNet convolution (MBConv) structure of the EfficientNet model. Additionally, a channel and spatial attention module (CBAM) is integrated before the final layer, accelerating model training and emphasizing nuanced features.

Results: The enhanced model, along with comparisons to VGGNet16, ResNet50, and DenseNet169, was subjected to ablation experiments. Our method achieved the highest classification accuracy of 95.12% on the cabbage vegetable image test set, outperforming VGGNet16, ResNet50, and DenseNet169 by 8.34%, 7%, and 4.29%, respectively. Notably, the proposed method effectively reduced the model's parameter count.

Discussion: Our experimental results highlight the effectiveness of the deep learning approach in improving vegetable quality grading accuracy. The superior performance of the enhanced EfficientNet model underscores its potential for advancing the field, achieving both high classification accuracy and parameter efficiency. We hope this aligns with your expectations. If there are further adjustments or clarifications needed, please let us know.

KEYWORDS

deep learning, vegetables, vegetable quality grading, EfficientNet network, attention module

1 Introduction

Due to the ongoing progress of economic globalization, there is a continual rise in both the variety and trading volume of agricultural products. Consequently, sales terminals have undergone significant historical transformations. The majority of companies operating in the vegetable industry refrigerate, store, and transport vegetables in a unified manner, lacking an explicit focus on the final consumer. They have a weak awareness of product grading and frequently employ traditional grading sorting methods, including relying on human labor for

sorting. This not only entails significant labor expenses but also yields vegetables of diverse quality, leading to diminished overall quality that could otherwise command a favorable market price. Additionally, vegetables with the potential for higher market value are acquired and packaged at a reduced cost, which directly affects the overall sales price and is not suitable for large-scale production. In comparison with traditional manual detection, recognition, and classification techniques, utilizing computer vision for image recognition, detection, and classification can not only enhance efficiency but improve accuracy as well. Currently, computer vision technology is widely employed in the classification of vegetables and fruits, the identification of plant and crop pests, and the identification of incomplete tablets, which can rapidly locate and identify the required features in detection; this achieves more efficient and economical extraction. The exploration of computer vision technology for assessing the visual quality of agricultural products has been conducted during the early stages of production, producing substantial outcomes. The primary emphasis has been on the examination of grains, dried fruits, fruits, eggs, and similar items. In recent years, with the substantial breakthroughs in deep learning technology in the field of image recognition, convolutional neural network models represented by VGGNet, GoogleNet, ResNet, etc., have not only achieved significant accomplishments (attained in extensive computer vision challenges) but have also been implemented by numerous scholars in the identification and categorization of vegetables and fruits, as well as the recognition of crop diseases and other domains. This has led to commendable outcomes in recognition accuracy. This also provides fresh ideas and the theoretical feasibility for vegetable image recognition methods. Consequently, to reduce the manpower, material resources, and costs required for classifying vegetable quality grades, this paper proposes a vegetable quality grading method on the basis of deep learning, establishes a vegetable grading image dataset, and subsequently proposes an improved EfficientNet model (CA-EfficientNet-CBAM) for vegetable quality grading. This results in savings in manpower and material resources, thereby reducing labor costs, enhancing vegetable grading performance, and expediting the speed of vegetable grading.

2 Literature review

The fruit and vegetable image classification technology process is predominantly divided into four steps. Step 1 involves inputting the image into the network model. Subsequently, in step 2, the input image undergoes preprocessing to extract more accurate and relevant features. Following this, step 3 focuses on the classification of the preprocessed image based on the extracted features. The classification of fruits and vegetables has progressed from approaches based on machine learning (Kurtulmuş and Ünal, 2015) to those based on deep learning (Latha et al., 2016; de Jesús Rubio, 2017; Pan et al., 2017) over time. Additionally, computer vision has substantially contributed to the field, particularly in the application of color sorting and grading for fruits and vegetables, which serves as a primary method to maintain product quality and increase overall value (Sun, 2000; Kondo, 2010; Patel et al., 2012; George, 2015; Xiao et al., 2015; Luo et al., 2021). The aforementioned developments are implemented in step 4, where computer vision methods are utilized to sort and grade colors.

Vegetable image classification based on traditional image processing preprocesses vegetable images and subsequently performs feature selection to classify and grade vegetables through color features, texture features, geometric features, etc. (Huang et al., 2023). Moreover, in 1996, Bolle et al. (1996) extracted the color, texture, and other features of vegetables to classify vegetable images. Nevertheless, when extracting features, external light easily interferes with this system, which affects the accuracy of recognition. Moreover, in 2010, Rocha et al. (2010) described the automatic classification of fruits and vegetables from images utilizing histograms, colors, and shape descriptions consistent with unsupervised learning methods. A 10-color model for defect detection and a rapid grayscale interception, with a segmentation threshold method to extract dark portions of a potato's surface, was both proposed by Li et al. (2010) in the same year. In 2012, to form the feature vector, Danti et al. (2012) first cropped and resized the images, subsequently extracted the mean and range of the hue and saturation channels of the hue, saturation, and value (HSV) image, and employed a backpropagation neural network (BPNN) classifier to process 10 types of leafy vegetables. In terms of classification, the success rate stands at 96.40%. In the same year, Suresha et al. (2012) utilized watershed segmentation to extract regions of interest as preprocessing and decision tree classifiers for training and classification. By employing texture measures in an red, green, and blue (RGB) color space, a dataset comprising eight distinct vegetables was acquired, achieving a classification accuracy of 95%. In 2015, Dubey and Jalal (2015) extracted different color and texture features after segmenting images and combining them. Experiments have demonstrated that when combining both yields, better recognition results can be obtained compared to utilizing separate color and texture features. In the same year, Madgi et al. (2015) proposed a vegetable classification method on the basis of RGB color and local binary pattern texture features.

In summary, most vegetable classification methods based on traditional image processing will extract features including color, texture, shape, etc. to detect and classify vegetables after preprocessing the image. Moreover, linear classifiers and K-nearest neighbor (KNN) classifiers are utilized in traditional classification. Due to the fact that the classifier needs to extract a large number of features to achieve optimal training results, during operation, it consumes significant memory and entails prolonged calculation times, thereby constraining the method's development and accuracy. Despite the fact that the utilization of machine learning methods for vegetable image recognition and classification can enhance the accuracy of classification, this method cannot be effectively applied in the recognition and classification of distinct vegetables.

The concept of deep learning (Wang et al., 2016; Gao et al., 2017; Lee et al., 2017) originated from artificial neural networks and has demonstrated excellent performance in feature learning and expression, which combines low-level features to form more abstract high-level features, thus discovering the distribution characteristics of data and enhancing image quality and recognition accuracy (Yang et al., 2022; Xu et al., 2023; Zhang et al., 2023). In 2020, Raikar et al. (2020) classified and graded okra fingers by comparing three models (AlexNet, GoogLeNet, and ResNet50) and employed transfer learning to train the network. Despite the fact that ResNet50 consumed the most training time, its accuracy was much higher than the other models. Gill et al. (2022) employed convolutional neural networks, recurrent neural networks, and long short-term memory to develop a

fruit image recognition system with multiple models. Moreover, the CNN extracted image features through different convolution layers, employed an recurrent neural network (RNN) to mark different features, and finally utilized long short-term memory (LSTM) to classify the optimal features extracted. Experimental evidence has established that this classification technology surpasses image classification technologies employing CNNs, RNNs, and RNN-CNNs in isolation. In a study by Li and Rai (2020), a deep convolutional neural network was introduced for the semantic segmentation of crops from a 3D perspective. This approach was designed to achieve efficient feature learning and object-based segmentation of crop plant objects within point clouds. Moreover, the experimental results indicate that eggplant and plant-level crop identification accuracy of cabbage is up to approximately 90%.

Ashtiani et al. (2021) developed a model that can accurately identify the maturity stage of mulberry trees by applying deep learning technology. Furthermore, the model utilizes deep learning algorithms for image recognition and analysis to differentiate the various stages of mulberry tree maturity. By extensively training and validating a substantial volume of mulberry tree image data, the researchers attained favorable outcomes. This investigation introduces a novel and effective approach for detecting mulberry tree maturity, holding significant potential for practical applications.

Similarly, Javanmardi et al. (2021) developed a computer vision classification system that can accurately classify distinct corn species by utilizing a deep learning model. By harnessing the image classification and feature extraction capabilities inherent in deep convolutional neural networks, this system attains precision in the classification of corn species. This accomplishment is realized through the extensive training and validation of a substantial dataset of corn seed images. Consequently, this research introduces an innovative and effective computer vision methodology for corn species classification, holding significant promise for diverse applications.

EfficientNet is a new convolutional neural network. In comparison with the previous convolutional neural network, this network uniformly scales the depth, width, and resolution of the network by setting fixed-scale scaling factors, with its high parameter efficiency and speed (Wang et al., 2021) being well-known. The efficiency of garbage classification was substantially improved; Jaisakthi et al. (2023) utilized the EfficientNet architecture based on transfer learning technology and also employed the Ranger optimizer to classify skin lesions in dermoscopic images. The Ranger optimizer was employed to classify EfficientNet and was optimized and fine-tuned to achieve an accuracy of 96.81%.

Based on the above research status both in China and abroad and following much of the literature on vegetable classification, this paper will discuss the employment of deep learning to study the vegetable quality grading problem and the design of a vegetable dish quality grading model based on deep learning. In the first place, it will address the needs of vegetable quality grading; we developed a dataset for research on quality grading. Second, we propose a vegetable dish quality grading algorithm, which combines the CA attention mechanism and the improved EfficientNet network fused with the CBAM module. The enhanced network model directs increased attention to subtle features, enabling the detection and localization of pertinent local information. The approach introduced in this manuscript not only enhances

accuracy but also diminishes the number of model parameters, presenting innovative methodologies and perspectives for research on vegetable classification and grading.

3 Research design

3.1 Co-ordinate attention

The squeeze-and-excitation network (SENet) was proposed by Hu et al. (2018) and won first place in the previous ImageNet2017 competition's classification task. Convolution typically focuses on the fusion of spatial scale information. By introducing an attention mechanism, SENet focuses on the connections between different channels to comprehend the importance of the characteristics of each channel. Furthermore, the innovation of the SENet network is to focus on the relationship between channels, with the expectation that the model will learn the significance of various channel features automatically.

The SE module calculates channel attention through 2D global pooling and achieves excellent performance at low computational cost, and yet this module does not consider the significance of position-related details. Moreover, the CA module (Hou et al., 2021) (co-ordinate attention) is an efficient lightweight attention module; the presence of this module enables the network to consider a more expansive region while maintaining a low computational cost, and it can be inserted arbitrarily. In each convolutional neural network, the feature expression ability of the network model is improved. The proposition of this module aptly addresses the challenge posed by employing two separate one-dimensional global pooling operations on input features in both the horizontal and vertical co-ordinate directions. This is achieved by utilizing the dual spatial ranges inherent in the pooling kernel. It is encoded and aggregated into a pair of direction-aware feature maps. In this pair of direction-aware feature maps, there are long-distance dependencies and precise position information along distinct spatial direction features.

The encoding operation of the co-ordinate attention mechanism is shown in Figure 1. First, the global pooling is decomposed, and the features are aggregated along two directions and converted into one-dimensional feature encoding. The mathematical formula for decomposition is illustrated in Eq. (1), where H and W represent the height and width of the feature map, respectively, and c indicates the number of channels. Subsequently, they are spliced and transformed by utilizing a convolution kernel with a convolution kernel size of 1×1 , as illustrated in Eq. (2). In Figure 1, r is the scaling factor. Subsequently, f is divided into two separate tensors along the dimensions of the space, the convolution operation is performed on them with a convolution kernel size of 1×1 , and the final feature map is output g^h, g^w . Finally, the output of the attention co-ordinate module is revealed in Eq. (3).

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_c(i,j) \quad (1)$$

$$f = \delta \left(F_1 \left[z^h, z^w \right] \right) \quad (2)$$

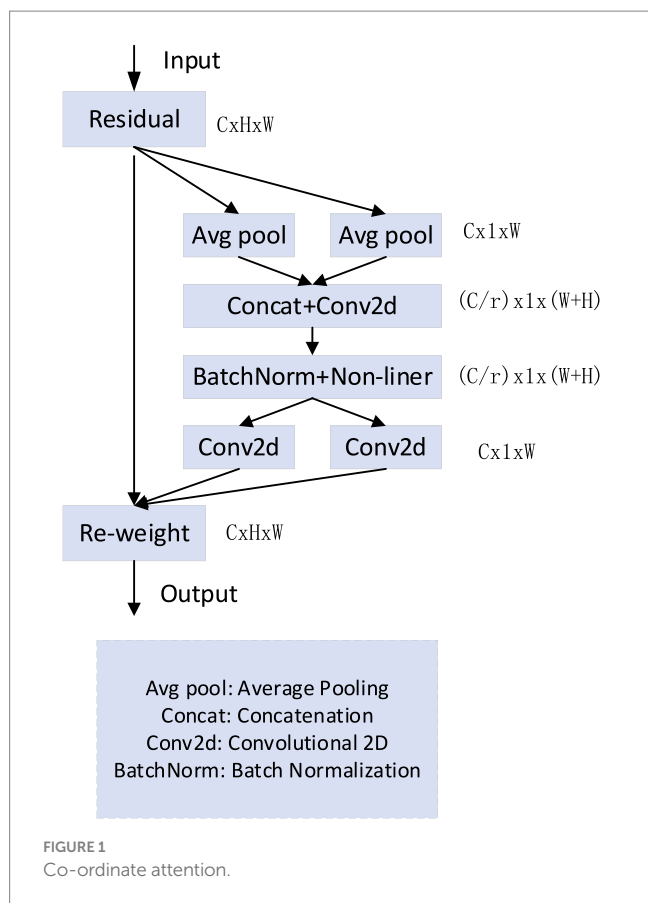


FIGURE 1 Co-ordinate attention.

TABLE 1 EfficientNet.

Stage	Operator	Resolution	Channels
1	Conv 3 × 3	224 × 224	32
2	MBCConv 3 × 3	112 × 112	16
3	MBCConv 3 × 3	112 × 112	24
4	MBCConv 5 × 5	56 × 56	40
5	MBCConv 3 × 3	28 × 28	80
6	MBCConv 5 × 5	14 × 14	112
7	MBCConv 5 × 5	14 × 14	192
8	MBCConv 3 × 3	7 × 7	320
9	Conv 1 × 1 & Pooling & FC	7 × 7	1,280

$$y_c(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j) \tag{3}$$

3.2 Improved MBCConv

In traditional neural networks, one of three methods (increase the depth of the network, modify the number of layers utilized in feature extraction, and increase the resolution of the input image) is typically employed to improve the accuracy and efficiency of the model. Nevertheless, as the depth of the network deepens, the gradient

problem of disappearance also appears. An augmentation in image resolution will correspondingly result in a surge in the computational workload of the model, albeit at the expense of a decline in accuracy. Tan and Quoc (2019) proposed the EfficientNet model, which adjusts the three dimensions of depth, width, and resolution of the image to enhance the accuracy of the model.

The architecture of the EfficientNet network is revealed in Table 1. The table indicates that the network is partitioned into nine distinct stages. The first stage has a convolution kernel size of 3 × 3 and a convolution layer with a stride of 2, and it contains batch normalization (BN) as well as Swish activation functions. The second-to-eighth stages are repeatedly stacked MBCConv structures. The ninth stage is composed of a convolution layer with a convolution kernel size of 1 × 1, an average pooling layer, and a fully connected layer. The depth, width, and resolution of the EfficientNet network are distinct from other networks that merely adjust one of them. By establishing a constant scale scaling factor, the network operates more efficiently in a three-dimensional environment due to uniform scaling.

In Table 1, FC represents “fully connected.” The backbone network of this model utilizes the mobile inverted bottleneck convolution (MBCConv) structure in the MobileNet V2 network. The SE attention mechanism is employed in the original MBCConv. In the EfficientNet network, the attention is to each lightweight inverted bottleneck convolution kernel. The mechanism is the SE module; however, this module disregards the location information of defects in vegetable quality grading and only considers the information encoding between channels. The defect information will directly influence the structure of plateau summer vegetable quality grading. In order to address this problem, SE is replaced by the CA attention mechanism in this paper. The replaced model is indicated in Figure 2. Moreover, the MBCConv structure includes two ordinary convolution layers with a convolution kernel size of 1 × 1 and a k × k depthwise conv, in which k × k has two structures: 3 × 3 and 5 × 5, a co-ordinate attention module, and a dropout layer. The two convolutional layers, specifically dimensionality reduction and augmentation, perform distinct purposes.

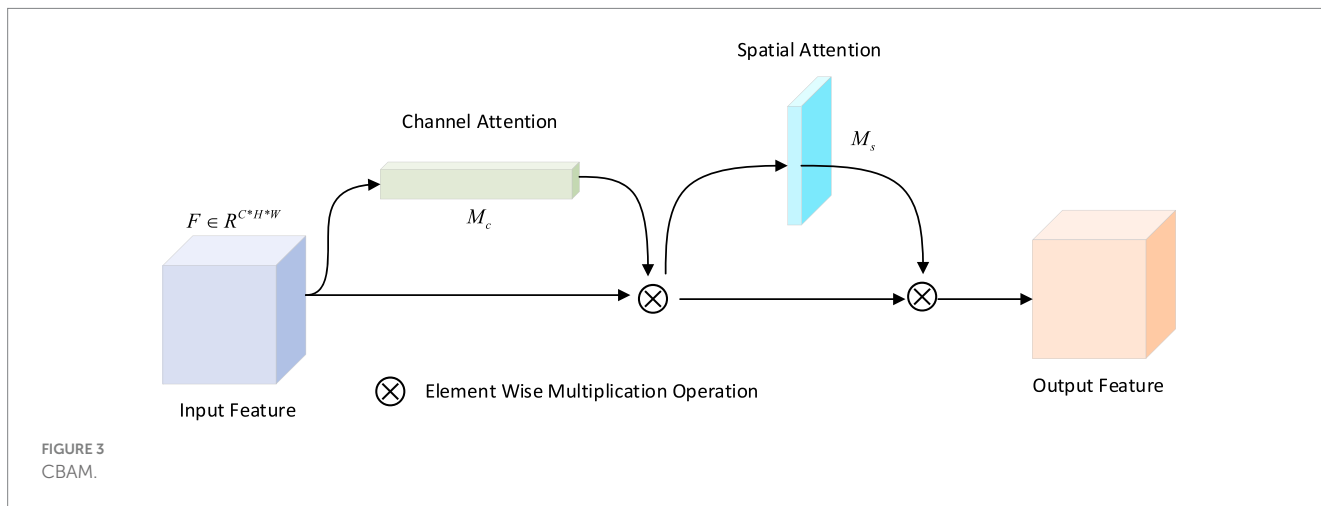
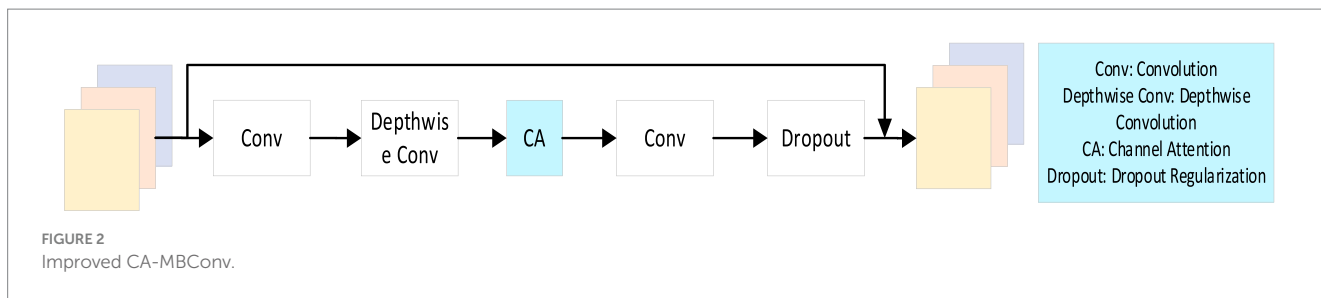
3.3 Convolutional block attention module

The CBAM (Woo et al., 2018) is a lightweight attention mechanism module. Two modules, namely the spatial attention module and the channel attention module, are utilized to process features. The channel attention mechanism can remove redundant feature information, and the spatial attention mechanism can remove irrelevant background information; its structure is shown in Figure 3.

The channel attention module forwards the input feature map and executes an element-wise summation operation on the feature output from the multi-layer perceptron, producing the ultimate channel attention feature map. We multiply it with the input feature map to generate the input features needed by the spatial attention module, as illustrated in Eq. (4):

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F))) + \text{MLP}(\text{MaxPool}(F)) \tag{4}$$

The spatial attention module utilizes the feature map output by using the previously stated channel attention as the input feature



map for this. This feature and the input to this module are multiplied to produce the final resulting feature map, as demonstrated in Eq. (5):

$$M_s(F) = \sigma\left(f^{7 \times 7}\left(\left[\text{AvgPool}(F); \text{MaxPool}(F)\right]\right)\right) \quad (5)$$

Among them, F is the input feature, $F \in R^{C*H*W}$ σ represents the activation function, $f^{7 \times 7}$ denotes the 7×7 convolution, $\text{AvgPool}(F)$ $\text{MaxPool}(F)$ are the average pooling feature and the maximum pooling feature, respectively. M_c represents the channel attention features, M_s represents the spatial attention features, and MLP represents the fully connected layers.

3.4 CA-EfficientNet-CBAM

In order to enhance the original network’s focus on nuanced features, detect and pinpoint local pertinent information, and improve the precision of analyzing similar species, this paper first replaces MBConv in the original EfficientNet with CA-MBConv after fusing the co-ordinate attention mechanism (CA). Second, the CBAM attention module is added prior to the last layer of the network, and the improved EfficientNet is indicated in Figure 4. After the input vegetable pictures are passed through the stacked CA-MBConv module, the output is employed as the input of the spatial attention module. Finally, the aggregate network module is derived by weighting the results acquired by the two modules.

4 Experiment

4.1 Dataset

Presently, there is a scarcity of publicly available datasets for vegetable quality grading. In addressing this issue, a standardized dataset incorporating vegetable classification and grading was constructed. Original images of vegetables were gathered using dedicated equipment to alleviate the current dataset deficiency in vegetable quality grading. In order to ensure the authenticity of the experiment, the data collected in this research came from supermarkets. We mainly collected data on six types of vegetables: lettuce, broccoli, tomatoes, garlic, bitter melon, and Chinese cabbage. We took 600 pictures of each type in the supermarket; the result is dataset A, which has a total of 3,600 photos.

As demonstrated in Table 2, in accordance with the appearance of vegetables, including firmness, size, tenderness, disease and insect infection, etc., each type of vegetable is divided into three levels, namely, special-grade vegetables, first-grade vegetables, and second-grade vegetables. Moreover, a detailed basis of the classification predominantly includes factors such as the size and visual appearance of the vegetables, the integrity and firmness of their outer bodies, the presence of surface defects, any indications of damage from pests and diseases, as well as mechanical damage. Additionally, the assessment considers the firmness and whiteness of the flower pattern, among other relevant attributes. Some experimental data are demonstrated in Figure 5. There are 200 special-grade vegetables, 200 first-grade vegetables, and 200 second-grade vegetables for each vegetable.

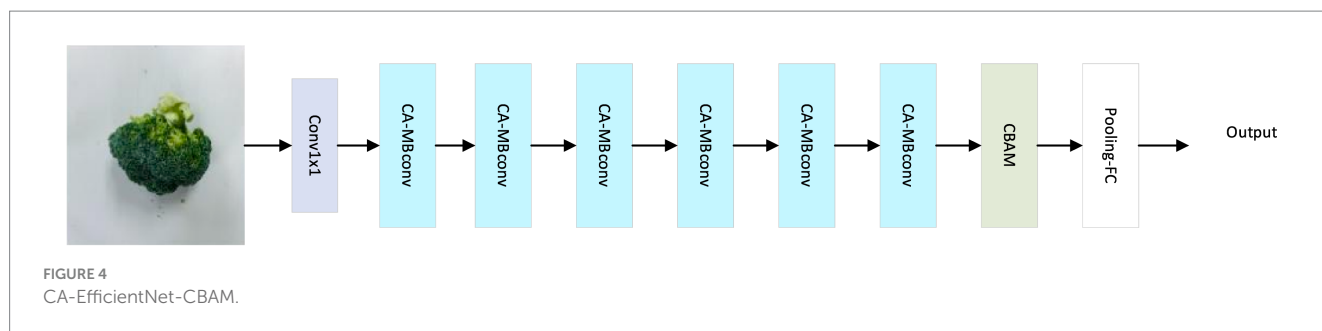


TABLE 2 Vegetable quality grading standards.

Special-grade vegetables	First-grade vegetables	Second-grade vegetables
The vegetables are neat and consistent in appearance, free from pests and diseases, and in the absence of mechanical damage. The vegetables look fresh	The vegetables are neat, with a few inconsistencies in the appearance of the vegetables, a small amount of damage, and no amount of damage from pests and diseases, and the surface of the vegetables is fresh	The vegetables exhibit irregularities in their appearance, featuring instances of disarray, substantial inconsistencies, considerable damage, minor pest and disease-related impairments, as well as instances where portions of the vegetable surfaces appear stale

Among them, 480 images of each type of vegetable were selected at random as the training set and 120 images as the test set.

4.2 Experimental details

Since the number of vegetable images collected in this experiment is insufficient, this paper employs the cross-validation method to divide the dataset to enhance the generalization ability of the model. Moreover, the performance of the deep learning model needs to be specifically quantified, compared, and analyzed through distinct indicators to verify the performance of the model. This study employs the improved EfficientNet network to grade the quality of vegetables. The performance of the model in the experiment was assessed and identified using accuracy and F_1 values. The objective was to compare the enhanced model to the original model and ascertain the improved model's limitations.

The experimental training was conducted through the AutoDL platform, which is a cloud GPU deep learning environment rental platform that is rich in resources and extremely efficient. Furthermore, the server employed is equipped with GPU: RTX 3090 (24 GB), CPU: 15 VCPUAMD EPYC 7543 32-Core Processor, and the environment for deep learning configuration is PyTorch 1.10.0 and Python 3.8.

The performance of the deep learning model needs to be specifically quantified, compared, and evaluated using various indicators to verify the performance of the model. This paper evaluates and detects the performance of the model through accuracy and F_1 value scores in terms of the quality grading experiment of plateau summer vegetables by utilizing the improved EfficientNet network. Concurrently, this paper employs the dimensions of the model parameters to assess the algorithm's intricacy, juxtaposes the initial model with the enhanced version, and substantiates the constraints of the improved model.

The accuracy rate represents the proportion of the number of correctly predicted images classified by the model to the total number of images. Moreover, the precision rate indicates the proportion of the number of correctly predicted images to the total number of positive predictions, and the recall rate reflects the number of correctly

predicted images to the total number of images that are positive. Given the inherent negative correlation between the precision rate and recall rate, conflicts commonly arise. In order to reconcile this conflict, the F_1 value was introduced as a holistic evaluation metric. This metric is derived through the weighted harmonic average of the two aforementioned rates. The proximity of the F_1 value to 1 indicates a superior performance for the network model. The accuracy rate and F_1 value are presented as per Equations (6) and (7).

$$\text{Accuracy} = \frac{TP + TN}{N} \tag{6}$$

$$F_1 = \frac{2TP}{N + TP - TN} \tag{7}$$

Among them, TP is the number of true samples; that is, there are positive samples in the dataset, and the prediction result is, likewise, a positive sample. TN is the number of true negative samples; that is, the dataset comprises negative samples, and the corresponding prediction outcome is also a negative sample. N is the total number of samples in the dataset.

In order to verify the performance of the improved model, this section analyzes the model training results. For the vegetable quality grading model CA-EfficientNet-CBAM network proposed in this paper, the experimental settings in Table 3 were employed for model training.

VGGNet is a deep convolutional neural network developed by researchers from the University of Oxford and Google DeepMind and is employed for image classification. The convolutional layer within this network exhibits a notable distinction from other networks. The spatial resolution of its feature map demonstrates an inverse relationship with the number of channels. The former progressively decreases, while the latter increases monotonically. This characteristic facilitates improved input image data processing. Additionally, the network repetitively integrates convolutional layers to construct a convolutional layer group, thereby augmenting the receptive field's scope and enhancing the network model's learning and feature expression capabilities.

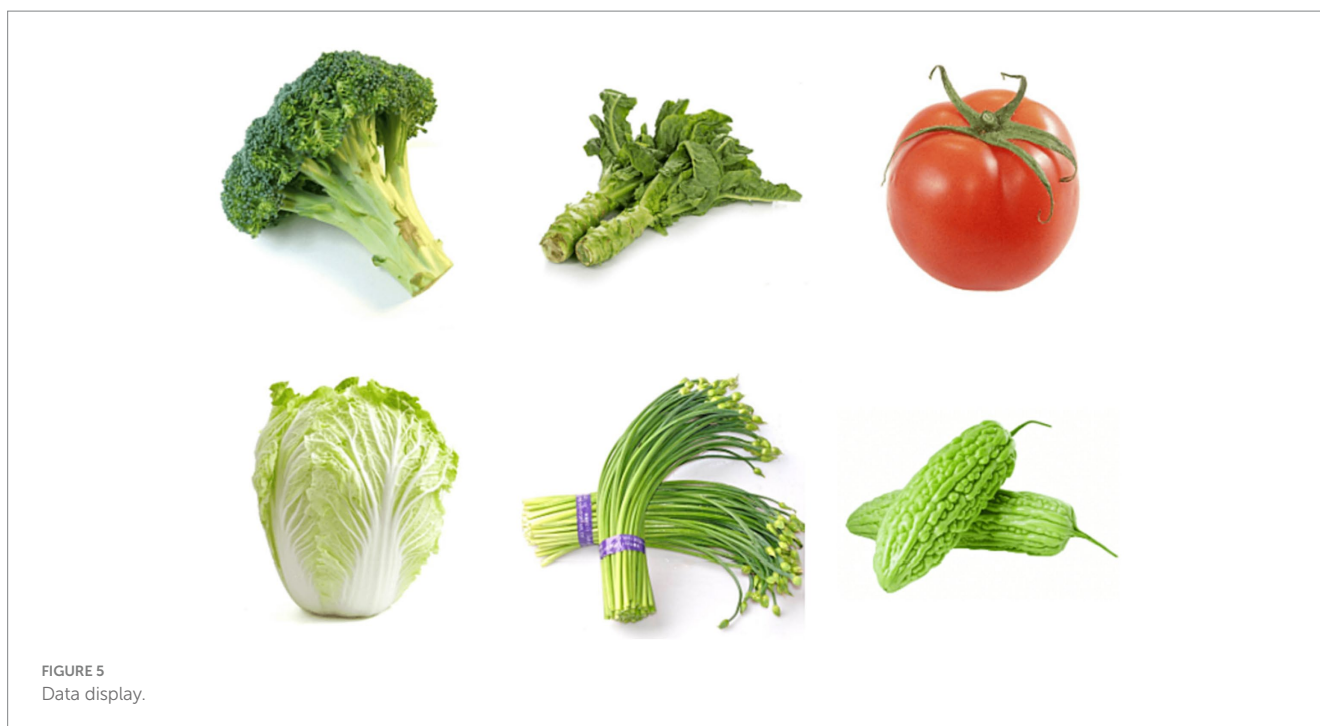


FIGURE 5
Data display.

TABLE 3 Training parameter settings.

Model parameters	Details
Batch size	16
Epoch	80
Initial learning rate	$6e-5$
Class num	3
Optimizer	Radom
Dropout	0.4

He et al. (2016) proposed the deep residual network ResNet. Compared with other convolutional neural networks, this network introduces identity mapping and calculates the residual to address the issue of degradation resulting from an excessive number of layers. The core of its model is establishing a “short-circuit connection” between the previous layer and the next layer, which assists in training a deeper network.

In comparison with ResNet, the DenseNet (Huang et al., 2017) model is independent of the deepening and broadening of the network structure and proposes a dense connection mechanism to achieve direct connections between levels, thereby improving network performance. To summarize, every layer acquires every feature map from the preceding layer. Consequently, all layers can establish direct connections with other layers that possess feature maps of equivalent sizes.

The comparative experimental results are shown in Table 4. The method mentioned in this paper possesses superior grading accuracy on the vegetable grading test set: 95.12%. Moreover, the improved EfficientNet model is better than VGGNet16, ResNet50, and DenseNet169 on the test set, and the accuracy rates increased by 8.34%, 6.67%, and 4.29%, respectively. In comparison to the above

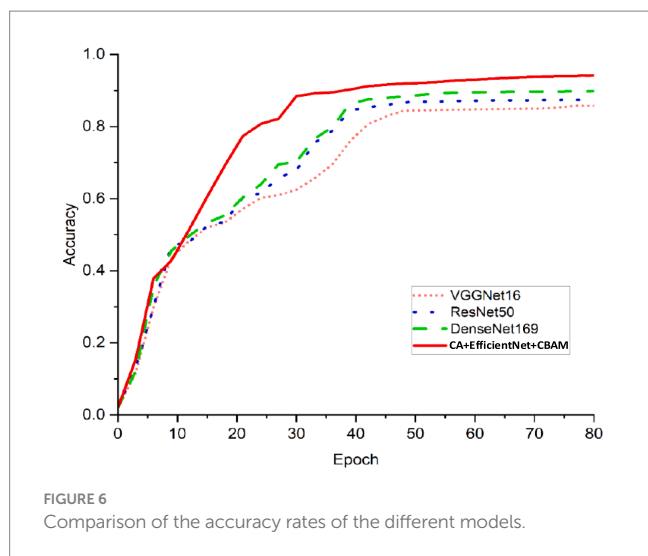
TABLE 4 Comparison of experimental results.

Model	ACC/%	F_1 /%	Parameter quantity/M
VGGNet16	86.78	80.24	47
ResNet50	88.45	82.26	42
DenseNet169	90.83	86.70	21
CA + EfficientNet + CBAM	95.12	92.10	5

three networks that enhance classification accuracy by increasing network depth, the improved model in this paper integrates the advantages of the lightweight module, CA module, CBAM module, and EfficientNet to reduce the amount of calculation and avoid excessive work during the fitting phenomenon, and efforts are made to preserve the characteristics of the input image to the greatest extent possible. In comparison to the three aforementioned classic convolutional neural networks, the approach advocated in this paper not only enhances accuracy but also mitigates the volume of model parameters.

The observation is evident from Figure 6: the accuracy of the four models increases with the number of iterations until it levels off. When the number of iterations reaches approximately 30, it becomes evident that the enhanced EfficientNet model exhibits the most rapid increase in accuracy and achieves significantly faster convergence compared to the other two models. However, as the number of iterations approaches 50, the accuracy of all four models tends to stabilize, showing similar performance levels.

For the purpose of further verifying the effectiveness of the improved model, this paper conducted four ablation experiments. Among them, Experiment 1 employed the original EfficientNet model to be trained on the vegetable quality grading dataset. Experiment 2 is based on the original EfficientNet model,



substituting the SE module in the MBConv structure with the CA module. Moreover, Experiment 3 was used to build a lightweight model by adding the lightweight module CBAM to the last layer of the network based on the original EfficientNet. Additionally, Experiment 4 is based on the original EfficientNet; that is, replacing the SE module in the MBConv structure with the CA module and adding a lightweight module (CBAM) to the last layer of the network to form the final improved EfficientNet model (refer to Table 5).

5 Conclusion

Due to the fact that the majority of vegetable quality grading utilizes traditional manual methods, this paper first gathered and produced six vegetable quality grading datasets and proposed an improved CA-EfficientNet-CBAM model for vegetable quality grading. First, the improved model was trained on six vegetable grading datasets and compared with the original model to test the grading effect of the improved model. Subsequently, the model was compared with the VGGNet16, ResNet50, and DenseNet169 network models on the vegetable grading dataset, and ultimately, an ablation experiment was conducted. Both the comparison experiment and the ablation experiment validated the viability and efficacy of the improved model.

References

- Ashtiani, S. H. M., Javanmardi, S., Jahanbanifard, M., Martynenko, A., and Verbeek, F. J. (2021). Detection of mulberry ripeness stages using deep learning models. *IEEE Access* 9, 100380–100394. doi: 10.1109/ACCESS.2021.3096550
- Bolle, R. M., Connell, J. H., Haas, N., Mohan, R., and Taubin, G. (1996). VeggieVision: a produce recognition system Proceedings Third IEEE Workshop on Applications of Computer Vision, 244–251.
- Danti, A., Manohar, M., and Basavaraj, S. A. (2012). Mean and range color features based identification of common Indian leafy vegetables. *Int. J. Image Process. Pattern Recognit.* 5, 151–160.
- de Jesús Rubio, J. (2017). A method with neural networks for the classification of fruits and vegetables. *Soft. Comput.* 21, 7207–7220. doi: 10.1007/s00500-016-2263-2
- Dubey, S. R., and Jalal, A. S. (2015). Fruit and vegetable recognition by fusing color and texture features of the image using machine learning. *Int. J. Appl. Pattern Recognit.* 2, 160–181. doi: 10.1504/IJAPR.2015.069538

TABLE 5 Ablation experiment results.

Improvement strategy	Accuracy/%	Parameter quantity/M
EfficientNet	90.92	7.64
CA + EfficientNet	92.76	4.65
EfficientNet + CBAM	93.24	4.66
CA + EfficientNet + CBAM	94.23	4.67

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

JW: Conceptualization, Formal analysis, Methodology, Writing – original draft. JH: Data curation, Formal analysis, Methodology, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Gao, X. H. W., Hui, R., and Tian, Z. M. (2017). Classification of CT brain images based on deep learning networks. *Comput. Methods Prog. Biomed.* 138, 49–56. doi: 10.1016/j.cmpb.2016.10.007

- George, M. (2015). Multiple fruit and vegetable sorting system using machine vision. *Int. J. Adv. Comput. Sci. Appl.* 6:2. doi: 10.4172/0976-4860.1000142

- Gill, H. S., Khalaf, O. I., Alotaibi, Y., Alghamdi, S., and Alassery, F. (2022). “Multi-model CNN-RNN-LSTM based fruit recognition and classification” in *Intelligent automation and soft computing*, Taylor & Francis, American Tech Science Press. 637–650.

- He, K. M., Zhang, X. Y., Ren, S. Q., and Sun, J. (2016). Deep residual learning for image recognition Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

- Hou, Q. B., Zhou, D. Q., and Feng, J. S. (2021). “Coordinate attention for efficient mobile network design.” 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, 13708–13717.

- Hu, J., Li, S., and Gang, S. (2018). "Squeeze-and-excitation networks." 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 7132–7141.
- Huang, G., Liu, Z., Maaten, L. V. D., and Weinberger, K. Q. (2017). Densely connected convolutional networks Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
- Huang, W., Wang, X., Zhang, J., Xia, J., and Zhang, X. (2023). Improvement of blueberry freshness prediction based on machine learning and multi-source sensing in the cold chain logistics. *Food Control* 145:109496. doi: 10.1016/j.foodcont.2022.109496
- Jaisakthi, S. M., Aravindan, C., and Appavu, R. (2023). Classification of skin cancer from dermoscopic images using deep neural network architectures. *Multimed. Tools Appl.* 82, 15763–15778. doi: 10.1007/s11042-022-13847-3
- Javanmardi, S., Ashtiani, S. H. M., Verbeek, F. J., and Martynenko, A. (2021). Computer-vision classification of corn seed varieties using deep convolutional neural network. *J. Stored Prod. Res.* 92:101800. doi: 10.1016/j.jspr.2021.101800
- Kondo, N. (2010). Automation on fruit and vegetable grading system and food traceability. *Trends Food Sci. Technol.* 21, 145–152. doi: 10.1016/j.tifs.2009.09.002
- Kurtulmuş, F., and Ünal, H. (2015). Discriminating rapeseed varieties using computer vision and machine learning. *Expert Syst. Appl.* 42, 1880–1891. doi: 10.1016/j.eswa.2014.10.003
- Latha, D., Jacob Vetha Raj, Y., and Mohamed, S. M. (2016). Multilevel approach of CBIR techniques for vegetable classification using hybrid image features. *ICTACT J. Image Video Process.* 6, 1174–1179. doi: 10.21917/ijivp.2016.0171
- Lee, S. H., Chan, C. S., Mayo, S. J., and Remagnino, P. (2017). How deep learning extracts and learns leaf features for plant classification. *Pattern Recognit.* 71, 1–13. doi: 10.1016/j.patcog.2017.05.015
- Li, J. W., Liao, G. P., Jin, J., and Yu, X. J. (2010). Method of potato external defects detection based on fast gray intercept threshold segmentation algorithm and ten-color model. *Trans. Chin. Soc. Agric. Eng.* 26, 236–242. doi: 10.3969/j.issn.1002-6819.2010.10.040
- Li, X., and Rai, L. (2020). "Apple leaf disease identification and classification using ResNet models." 2020 IEEE 3rd International Conference on Electronic Information and Communication Technology (ICEICT), Shenzhen, China, 738–742.
- Luo, J., Zhao, C., Chen, Q., and Li, G. (2021). Using deep belief network to construct the agricultural information system based on internet of things. *J. Supercomput.* 78, 379–405. doi: 10.1007/s11227-021-03898-y
- Madgi, M., Ajit, D., and Basavaraj, A. (2015). Combined RGB color and local binary pattern statistics features-based classification and identification of vegetable images. *Int. J. Appl. Pattern Recognit.* 2, 340–352. doi: 10.1504/IJAPR.2015.075947
- Pan, L., Sun, Y., Xiao, H., Gu, X., Hu, P., Wei, Y., et al. (2017). Hyperspectral imaging with different illumination patterns for the hollowness classification of white radish. *Postharvest Biol. Technol.* 126, 40–49. doi: 10.1016/j.postharvbio.2016.12.006
- Patel, K. K., Kar, A., Jha, S. N., and Khan, M. A. (2012). Machine vision system: a tool for quality inspection of food and agricultural products. *J. Food Sci. Technol.* 49, 123–141. doi: 10.1007/s13197-011-0321-4
- Raikar, M. M., Meena, S. M., Chaitra, K., Shantala, G., and Pratiksha, B. (2020). Classification and grading of okra-ladies finger using deep learning. *Procedia Comput. Sci.* 171, 2380–2389. doi: 10.1016/j.procs.2020.04.258
- Roche, A., Hauagge, D. C., Wainer, J., and Goldenstein, S. (2010). Automatic fruit and vegetable classification from images. *Comput. Electron. Agric.* 70, 96–104. doi: 10.1016/j.compag.2009.09.002
- Sun, D. W. (2000). Inspecting pizza topping percentage and distribution by a computer vision method. *J. Food Eng.* 44, 245–249. doi: 10.1016/S0260-8774(00)00024-8
- Suresha, M., Sandeep, K. K. S., and Shiva, K. G. (2012). Texture features and decision trees based vegetable classification IJCA Proceedings on National Conference on Advanced Computing and Communications, 975: 8878.
- Tan, M. X., and Quoc, L. (2019). "Efficientnet: rethinking model scaling for convolutional neural networks." International Conference on Machine Learning (PMLR).
- Wang, P., Li, W., Liu, S., Gao, Z., Tang, C., and Ogumbona, P. "Large-scale isolated gesture recognition using convolutional neural networks." 2016 23rd International Conference On Pattern Recognition (ICPR), Cancun, Mexico, (2016), 7–12.
- Wang, J., Liu, Q., Xie, H., Yang, Z., and Zhou, H. (2021). Boosted EfficientNet: detection of lymph node metastases in breast cancer using convolutional neural networks. *Cancers* 13:661. doi: 10.3390/cancers13040661
- Woo, S., Park, J., Lee, J. Y., and Kweon, I. S. (2018). CBAM: convolutional block attention module Proceedings of the European Conference on Computer Vision (ECCV).
- Xiao, Q. M., Niu, W. D., and Zhang, H. (2015). "Predicting fruit maturity stage dynamically based on fuzzy recognition and color feature." 2015 6th IEEE International Conference on Software Engineering and Service Science (ICSESS), 944–948.
- Xu, J., Yang, Z., Wang, Z., Li, J., and Zhang, X. (2023). Flexible sensing enabled packaging performance optimization system (FS-PPOS) for lamb loss reduction control in E-commerce supply chain. *Food Control* 145:109394. doi: 10.1016/j.foodcont.2022.109394
- Yang, Z., Xu, J., Yang, L., and Zhang, X. (2022). Optimized dynamic monitoring and quality management system for post-harvest matsutake of different preservation packaging in cold chain. *Foods* 11:2646. doi: 10.3390/foods11172646
- Zhang, H., Zuo, X., Sun, B., Wei, B., Fu, J., and Xiao, X. (2023). Fuzzy-PID-based atmosphere packaging gas distribution system for fresh food. *Appl. Sci.* 13:2674. doi: 10.3390/app13042674