CERASUS HUMILIS CULTIVARS IDENTIFICATION WITH SMALL-SAMPLE AND UNBALANCED DATASET BASED ON EFFICIENT NET-B0+RANGER NETWORKS
/
*基于 **EfficientNet-B0+Ranger** 的小样本不平衡数据集欧李品种识别*

Lili LI [1)]; Hua YANG [*1)]; Bin WANG [1)] [1]
[1)] College of Information Science and Engineering, Shanxi Agricultural University, Taigu 030800/China
Tel: 13466883256; E-mail: yanghua@sxau.edu.cn
DOI: https://doi.org/10.35633/inmateh-69-21

Keywords: *cerasus humilis cultivars, efficientNet-B0+Ranger, small-sample, unbalanced dataset, complex environment*

ABSTRACT
Because of the high similarity of leaves of different Cerasus humilis varieties, it is difficult to identify them with the naked eye. In this study, the leaves of four different Cerasus humilis varieties collected in the field were used as the research objects, and a new leaf recognition model based on the improved lightweight convolution neural network model EfficientNet-B0 was proposed. Firstly, the performance of the network models Efficientnet-B0 and ResNet50, GoogleNet, ShuffleNet, and MobileNetV3 were compared based on two different learning methods. Then, the influence of different optimizers on model recognition accuracy was compared based on the optimal model. Finally, different learning rates were used to optimize the optimal model. The results show that the recognition rate of the proposed Efficientnet-B0 +Ranger+0.0005 model was up to 86.9%, which was 2.23% higher than that of the original Efficientnet-B0 model. The results show that this method can effectively improve the recognition accuracy of Cerasus humilis auriculate leaves, which can provide a reference for the deployment of the leaf identification model of Cerasus humilis variety on the mobile terminal.

摘要
针对不同欧李品种叶片相似度高，用肉眼难以鉴别的问题。本研究以田间收集的4种不同欧李品种叶片为研究对象，提出一种基于改进轻量级卷积神经网络模型EfficientNet-B0的欧李品种叶片识别模型。首先，基于2种不同学习方式对比EfficientNet-B0与ResNet50, GoogleNet, ShuffleNet, MobileNetV3等网络模型的性能；然后，基于最优模型对比不同优化器对模型识别准确率的影响；最后采用不同学习率（0.0001, 0.0005, 0.001, 0.005, 0.05)对优选出的模型进行优化。结果表明该研究提出的EfficientNet-B0+Ranger+0.0005模型识别率达到86.9%，相比于改进前的EfficientNet-B0模型，其识别率提高了2.23%。研究结果表明，该方法可有效提高欧李品种叶片的识别准确率，可为移动端部署欧李品种叶片识别模型提供参考。

INTRODUCTION
Cerasus humilis (Bge.) Sok. belongs to the Rosaceae cherry genus shrub fruit tree. It is a unique fruit tree resource in China. Cerasus humilis fruit is rich in active calcium and easy to be absorbed by the human body, it is also known as the “fruit rich in calcium”. It has the ability of early fruiting, high yield, cold resistance, drought resistance, stubble resistance, wind and sand fixation, and soil erosion control. It is a strategic tree species for sustainable development *(Bai et al., 2015)*.

In recent years, with the increasing variety of Cerasus humilis cultivars, the planting area has gradually expanded. However, the pest control and management of different cultivars of Cerasus humilis are different. The identification of Cerasus humilis cultivars is mostly carried out by plant protection experts, which often has a high recognition rate, but it is time-consuming, labor-consuming, subjective, and with the given scientific and genetic advances, many of the hybrid Cerasus humilis cultivars cannot be identified by ordinary individuals and even gardening experts *(Ferentinos et al., 2018)*. So, computer vision and machine learning techniques are used to solve these problems. The images of leaves had been used by many researchers of computer vision to identify species of plants *(Kalyoncu et al., 2015; John et al., 2021).*

[1] *Lili Li, Lec. Ph.D.; Hua Yang**, *Prof. Ph.D.; Bin Wang, Ph.D.*

Leaf shape is the feature most commonly used to develop automated plant classification systems. Other than the shape *(Mouine et al., 2012)*, the leaf can provide some additional information such as textures *(Naresh et al., 2016)*, venation *(Charters et al., 2014)*, and colors *(Tan et al., 2018)*. Although there are many studies aimed to classify different plant species, they were based on the common fact that the leaves they used had different colors and shapes, which made the classification task easier. In contrast, the leaves of Cerasus humilis cultivars are generally of the same color and similar in shape, so using these methods to classify Cerasus humilis cultivars with leaf images used the method of computer vision is challenging and problematic.

Recently, deep learning methods especially the Convolutional Neural Network (CNN) had been employed for the automated extraction of features. It can extract more detailed information as compared to conventional machine learning techniques. And deep learning models have been well used in the field of agriculture, such as weed and seed detection, counting, plant disease identification, root segmentation, etc. The common deep learning classification methods based on the convolutional neural network (CNN) are AlexNet, VGGNet, GoogleInceptionNet, and ResNet.

Liu et al. (2018) proposed a ten-layer CNN for plant leaf classification on a leaf database of Flavia which contains 4,800 leaf images and 32 kinds of leaves, and the result showed that the proposed method achieved a high overall accuracy of 87.92%. *Beikmohammadi et al. (2020)* proposed a new method and compared it with three deep learning-based models (S-Leaf Net, W-Leaf Net, and P-Leaf Net). The tests were conducted on the two well-known datasets of leaf identification, i.e. MalayaKew (MK) and Flavia. The results showed that the proposed method obtained 99.81% and 99.67% of accuracy and outperformed all the other methods. Unfortunately, when transferring these methods from species identification to cultivar identification, most of them performed worse due to the high similarity of leaves among different cultivars.

Although the deep network model had achieved excellent performance in the plant leaf classification task. However, considering the real environment of crop growth, deploying trained deep network models on mobile phones or embedded devices still faces significant challenges. The main influencing factors include huge model parameters, long training time, and large storage space *(Li et al., 2019)*. Therefore, the current research tends to miniaturize and apply the model structure. Among them, *Sun et al. (2017)* proposed an improved convolutional neural network model that combines batch normalization (BN) and global pooling to train 21917 leaf images collected by PlantVillage. After only three training sessions, the recognition accuracy reached over 90%.

Guo et al. (2019) proposed a convolutional neural network image recognition method for plant leaves based on transfer learning, and the recognition accuracy achieved by AlexNet and InceptionV3 models reached 95.31% and 95.40%, respectively.

Liu et al. (2019) proposed two lightweight convolution neural network models for crop disease classification based on MobileNet and Inception V3, comprehensively considered the recognition accuracy, operation speed, model size, and other indicators to select the best, and realized the leaf detection of a mobile terminal. At present, most of the methods are based on public datasets, which contain simple backgrounds and many plant species, and there is little research on the identification of the plant cultivars, such as the author *Zhang et al. (2021)* proposed an automatic leaf image-based cultivar identification pipeline called MFCIS (Multi-feature Combined Cultivar Identification System), to identify 88 Sweet cherry cultivars and 100 Soybean cultivars and achieved an overall accuracy of 83.52% and 91.4%, respectively.

Liu et al. (2020) proposed an efficient and convenient method for the classification of 14 apple cultivars using a deep convolutional neural network and achieved an overall accuracy of 97.11%. Although these studies aimed to classify different plant cultivars the images they used were also simple backgrounds. However, there is little research on the identification of plant cultivars in the field environment.

Based on these problems, this paper used the classic lightweight CNN structure EfficientNet to identify 4 Cerasus humilis cultivars by using the leaf images collected in the natural environment. To deploy the trained model to the mobile terminal to establish an intelligent detection system, and realize the rapid identification of Cerasus humilis cultivars according to the leaves.

The remainder of this paper is organized as follows. In section 2 the datasets, data augmentations performed, and Cerasus humilis cultivars diagnosis model are presented, Section 3 presents the experimental setup as well as analyzes the experimental results provided by the identification approach to Cerasus humilis cultivars based on CNNs. In section 4 the significance of the identification of Cerasus humilis cultivars is discussed. Finally, this paper is concluded in Section 5.

MATERIALS AND METHODS

In this section, the procedure of identifying Cerasus humilis cultivars by using leaf is segregated into seven steps processes which showed in Fig. 1.

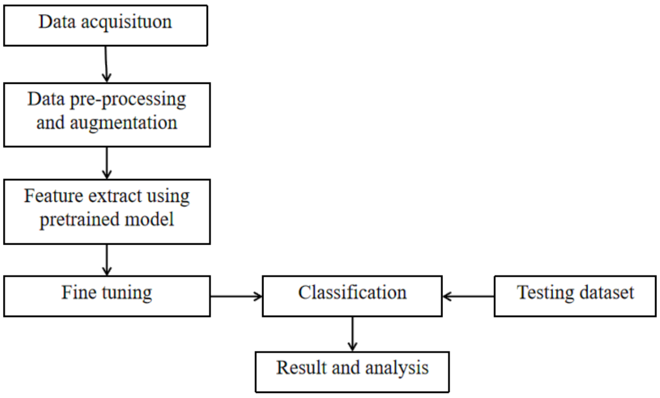


Fig. 1 - The procedure of the work

Dataset

The 4 Cerasus humilis cultivars leaf images used in this study were collected from the Cerasus humilis planting demonstration base in the national (Jinzhong) agricultural high-tech industry demonstration zone. The base covers an area of about 6000 square meters, with a total of 1000 Cerasus humilis trees, all of which are 4 years old. In the process of taking photos, it is also considered the influence of angle, illumination, occlusion, and other complex environments, to take into account the influence of as many environmental conditions as possible in the process of data acquisition. And the samples were labeled according to the cultivars of the Cerasus humilis plant, and the label name was the name of Cerasus humilis cultivars.

Under the natural environment conditions, multi-angle collections were carried out at different times. The mobile phone camera lens was used to take images 10~20 cm away from the Cerasus humilis leaf. The images dataset includes 4 Cerasus humilis cultivars, that is Nongda-4 Cerasus humilis leaves (403 images), Nongda-5 (476 images), Nongda-6 (581 images), and Nongda-7 (558 images). Examples of leaf images for all 4 Cerasus humilis cultivars are shown in Fig. 2.

a) Nongda-4 *b) Nongda-5* *c) Nongda-6* *d) Nongda-7*

Fig. 2 - Leaves of different varieties of Cerasus humilis

Fig. 2 reveals that the leaves of 4 Cerasus humilis cultivars are generally very simple and similar to each other, and they have an elliptical-to-ovate shape and small difference in length and width, with a sharp apex and round and blunt serrated edges. Due to these similarities, there are challenges in classifying Cerasus humilis cultivars.

Image Preprocessing and Augmentation

Image Preprocessing

Image preprocessing helps to improve data quality and enhance transform information features. The preprocessing of experimental data in this study includes redefinition of image size, pixel averaging, and normalization of each batch of samples. In this experiment, the image is redefined as 224×224×3. The purpose of image de-averaging is to standardize the image and remove the average brightness value of the image. The specific operation is as follows: for the given size redefined image, each pixel value of each sample subtracts the average pixel value of the whole training sample. In many cases, it is not sensitive to the brightness of the image but pays more attention to its content. For CNNs, data normalization is necessary for gradient descent, and it can also prevent gradient explosion, accelerate network convergence and further reduce the number of feature maps. The batch normalization method is used in this experiment, and each sample image in the batch needs to be normalized. The processing of the image x is shown in Equations (1) and (2):

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

$$\sigma = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu)^2 \tag{2}$$

Where x_i represent the value of the i-th pixel of the sample, n is the total number of pixels of the sample, μ represents the mean value, and σ represents the variance.

The normalization is shown in Equation (3):

$$\hat{x}_i = \frac{x_i - \mu}{\sqrt{\sigma^2 + \varepsilon}} \tag{3}$$

Where $\hat{x}_i$ is the normalized pixel value of the i-th pixel of the image x, ε is a small value greater than 0 to ensure that the denominator is greater than 0.

Image Augmentation

Image augmentation can increase the diversity of images, reduce the dependence of the model on some attributes, and improve the performance and generalization ability of the model. The following image augmentation technique was mainly used: random rotation angle (the maximum rotation angle is 30º), random horizontal or vertical translation (the maximum translation distance is 0.1× image width or height), random horizontal flip and other operations, and slight disturbance were introduced to realize data expansion, the missing pixels were filled with the adjacent pixels after image enhancement. In deep learning, the uneven distribution of samples will affect the accuracy. Finally, in order to meet the data balance of each category, after image augmentation, the image became Nongda-4 (3180 images), Nongda-5 (2555 images), Nongda-6 (3140 images), Nongda-7 (3095 images). Whereafter, according to the ratio of 8:1:1, it was randomly divided into the training set, validation set, and test set.

Convolutional Neural Network Models

Actually, for obtaining an optimal model, we tried several deep learning models. In this section, there will be simply introduced state-of-the-art CNNs such as ResNet50, GoogleNet, MobileNetV3, ShuffleNet, EfficientNet-B0 *(Tan et al., 2018)*, and improved EfficientNet-B0.

EfficientNet Model

The birth of CNNs promotes the development of deep learning, from the original convolutional layer, pooling layer, and full connection layer, which consists of simple network LeNet, AlexNet, and VGG-16 *(Charters et al., 2014)*, to ResNet, Inception, and GoogleNet.

In order to improve the accuracy and efficiency of the model, the current general method is to amplify the depth, width, and resolution of CNN. Previously, it was used to amplify one of the three dimensions alone. For example, ResNet18 to ResNet152 increased the network depth to improve the accuracy. Although it is possible to enlarge two or three dimensions arbitrarily, it needs tedious manual parameter adjustment and may produce a suboptimal accuracy and efficiency. The experimental study in reference showed that the three dimensions of depth, width, and resolution affect each other. How to balance these three dimensions is crucial. It is verified that such a balance can scale each dimension by simply using a set of constant ratios. On this basis, the latest network EfficientNet was proposed. In the model, a simple and efficient multi-dimensional hybrid scaling method was proposed, that is to find a set of fixed scaling factors that can take into account both speed and accuracy to scale the depth, width, and resolution of the network.

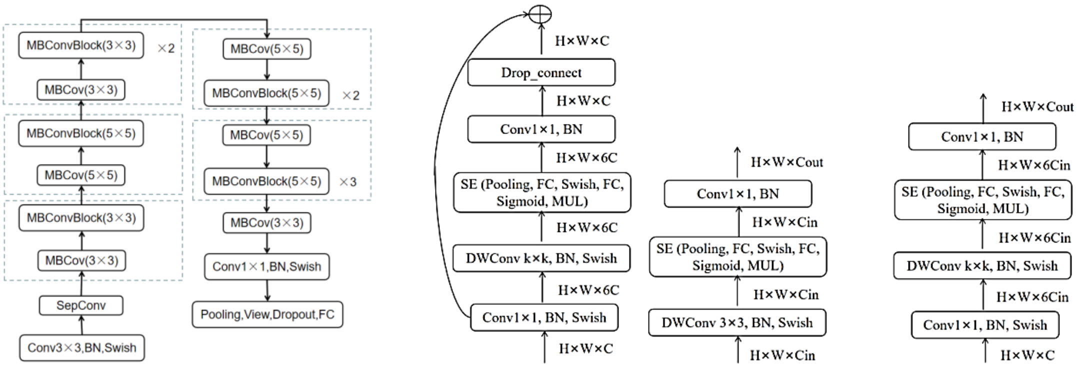
The EfficientNet group consists of 8 models between B0 and B7, and the EfficientNet-B0 network structure consists of 16 MBConv+2 Conv+1 Global average pooling + 1 FC classification layer. In addition, MBConv (mobile inverted bottleneck convolution) module is used in the network, and the attention idea of the SENet (Squeeze-and-Excitation Network) is introduced.

The overall structure of the EfficientNet-B0 was shown in Table 1, and the network structure can be seen in Figure 3. Where (a) is a representative model selected from Table 1; (b)-(d) are a few corresponding layer structures. MBCov denotes mobile inverted bottleneck Conv, DWCov denotes depthwise Conv, k3×3/k5×5 denotes kernel size, BN is batch normalization, and H×W×C denotes tensor shape (height, width, channel).

Table 1

The overall structure of the EfficientNet-B0

Stage	Operator	Resolution	#Channels	#Layers
i	$\hat{F}_i$	$\hat{H}_i \times \hat{W}_i$	$\hat{C}_i$	$\hat{L}_i$
1	Conv3×3	224×224	32	1
2	MBConv1,k3×3	112×112	16	1
3	MBConv6,k3×3	112×112	24	2
4	MBConv6,k5×5	56×56	40	2
5	MBConv6,k3×3	28×28	80	3
6	MBConv6,k5×5	14×14	112	3
7	MBConv6,k5×5	14×14	192	4
8	MBConv6,k3×3	7×7	320	1
9	Conv1×1&Pooling&FC	7×7	1280	1



a) The EfficientNet-B0 structure *b) MBConvBlock(k×k)* *c) SEPConv* *d) MBconv (k×k)*

Fig. 3 - The EfficientNet-B0 structure

Unlike other CNN models, EfficientNet uses a new activation function called Swish instead of the Rectifier Linear Unit (ReLU) activation function. And the EfficentNet-B0 contains 5.3M parameters, as it is suitable according to our resources and purpose.

Swish

The choice of activation functions in deep networks has a significant effect on the training dynamics and task performance. Currently, the most successful and widely-used activation function is the Rectified Linear Unit (ReLU). Although various alternatives to ReLU have been proposed, none have managed to replace it due to inconsistent gains. In 2017, *Ramachandran et al. (2017)* proposed to leverage automatic search techniques to discover new activation functions named Swish, that is, $f(x) = x \cdot sigmoid(\beta x)$. Where β is a hyperparameter.

The form of the Swish function is very simple, and almost a patchwork of Sigmoid and ReLU, with characteristics of smooth, non-monotonic, with no upper bound but a lower bound. The Swish tends to work better than ReLU on deeper models across several challenging datasets. And the simplicity of Swish and its similarity to ReLU make it easy for practitioners to replace ReLUs with Swish units in any neural network.

Improved EfficientNet-B0 Model

To further improve the accuracy of the EfficientNet model, we adopted the current state-of-the-art Ranger optimizer proposed by Dang. The Ranger optimizer combines the ideas of RAdam and LookAhead. Among them, RAdam (Rectified Adam) is a variant of Adam, which utilizes dynamic rectifiers to adjust Adam's adaptive momentum according to variance, and provides an automatic warm-up mechanism effectively customized based on the current dataset to ensure a solid first step in training. LookAhead was inspired by the latest developments in deep neural network loss surface understanding and provided a breakthrough in robust and stable exploration throughout the entire training period. The LookAhead optimizer reduces the need for extensive hyperparameter adjustment, while achieving faster convergence of different depth learning tasks with minimal computational overhead. Therefore, the combination of RAdam and LookAhead optimizers had high synergy and can provide the best improvement in deep learning optimization.

Metrics

In order to evaluate the results of the proposed method for the identification of Cerasus humilis cultivars, the average identification accuracy, and the model parameters were considered respectively.

Average Accuracy (AA)

Average Accuracy (AA) represents the proportion of currently classified predictions, it refers to the ratio of the predicted correct samples to all the observed samples. It is the most important metric index to evaluate the performance of the model, and is calculated as equation (4):

$$\mathrm{AA} = \frac{1}{n_s}\sum_{i=1}^{n_s}\frac{n_{ii}}{n_i} \tag{4}$$

Where n_s is the total number of sample categories (in this paper, the focus is on the four sample categories of Cerasus humilis cultivars, so $n_s = 4$), i is the label of the sample category, n_i is the total number of class i samples, n_{ii} is the total number of class i samples (that is, the number of correctly predicted samples in each type).

Model Parameters

With the improvement of the accuracy of the network models, the hardware resources required for the deployment and operation of the models are considered while improving the identification accuracy. Too complex network models occupy a lot of storage and computing resources, which limits their deployment and operation in the scenario of limited hardware resources. Therefore, the size of the model is also one of the key factors to be considered.

Transfer Learning

Deep learning usually requires a large number of training samples, and the models with small images are prone to over-fitting, thus reducing the generalization ability of the models. Deep learning involves initializing weights on pre trained networks on large labeled datasets (such as ImageNet datasets), rather than randomly initializing from existing sample sets.

However, transfer learning is to adapt the convolutional neural network model trained on a task to a new task through simple adjustment. And it can reduce the number of images, save time, and improve the accuracy of training models.

The cerasus humilis leaf dataset in this study is a small sample dataset. The transfer learning method will be used to fully train the pre trained models (such as MabileNetV3, ResNet50, GoogLeNet, ShuffleNet, and EfficientNet-B0) and apply them to the recognition of different kinds of leaf images.

Experiment Setup

The training and testing of the model were completed under the Cuda deep learning framework. Hardware environment of the platform: NVIDIA GTX 1650 GPU, 16G graphics memory, Intel Xeon Gold 6271C @2.60GHz CPU, 32GB of memory. Software environment: Python 3.7, Cuda10.1.

To avoid memory overflow, batch training methods were used to compare the EfficientNet-B0 model with Google Net, MobileNetV3, ShuffleNet, and ResNet50 on the training and validation sets. During the training and validation process, 16 images were set for each batch. In this paper, the number of iterations was set to 50, and the Categorical-cross entropy in Keras was used as the cost function. In order to solve the problem of gradient disappearance and explosion in the process of backpropagation, using batch normalization (BN) to normalize the characteristic map of each sample formed after the convolution layer can well solve the problem that the distribution of data in the middle layer changes during the training process of the model, accelerate convergence, improve accuracy, and reduce overfitting.

To accurately obtain the parameters of the training model, the optimal model is determined by observing the change of the loss function of the validation set after each iteration, and the model is applied to the test set to realize the prediction of leaf varieties.

RESULTS AND ANALYSES

The Compared Results of Different Transfer Learning Strategies

The main purpose of this study is to examine the success of EfficientNet-B0 deep learning architecture in the identification of Cerasus humilis cultivars and compare it with the performances of state-of-the-art CNN models in the literature. Three sets of comparative experiments of different learning methods, different optimizers, and different learning rates are carried out.

Two transfer learning strategies had been considered as follows. One is that instead of random initialization, the network weights are initialized by the pre-trained network model in the ImageNet data set and global fine-tuning is performed. Another method was that freeze the weights of all networks, except for the final fully connected layer. And use the pre trained model as a feature extractor (ResNet50, MobileNetV3, and EfficientNet-B0) to learn the corresponding features of the target dataset by fine-tuning the fully connected layer. In this section, the optimizer selected is Adam, and the learning rate was set to 0.0001.

The First Transfer Learning Strategy

In this section will be given the testing results of 5 kinds of deep learning models, including MobileNetV3, GoogleNet, ResNet50, ShuffleNet, and EfficientNet-B0, by using the transfer learning method of strategy one. Among them, ResNet50 and GoogleNet belong to traditional CNNs, and the other three belong to lightweight CNNs.

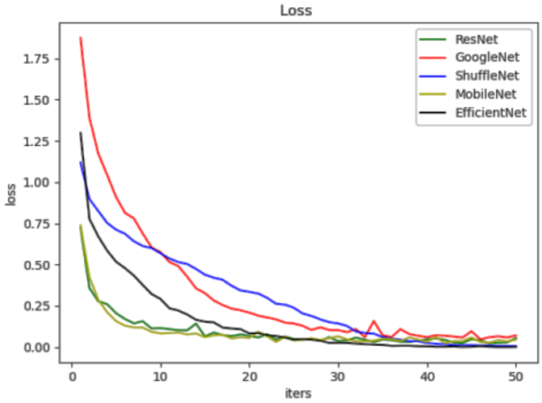
In this context, the average accuracy on the test set, and model parameters obtained by all models were given in Table 2 for the 4 Cerasus humilis cultivars leaf datasets.

Table 2

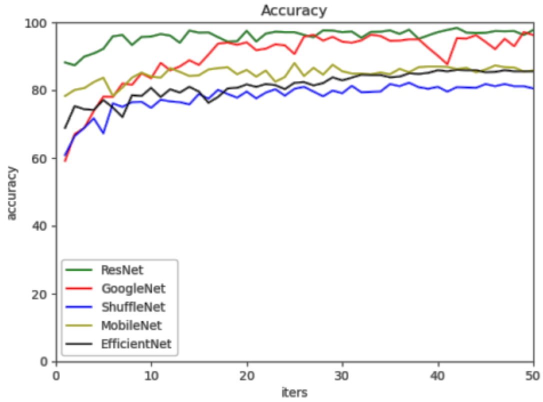
The comparison results of different models

Model	Model parameters	Average accuracy(%)
ResNet50	25,664,453	97.89
GoogleNet	10,318,655	95.72
MobileNetV3	4,232,837	84.29
ShuffleNet	1,274,909	75.12
EfficientNet-B0	4,055,976	84.67

The results showed that although the recognition accuracy of the traditional deep learning model ResNet50 and GoogleNet was higher than that of the lightweight model, their model size and average recognition time were also much larger than the other three lightweight models. The accuracy rates of the two lightweight models are all above 84.0%, except for ShuffleNet 75.12%, which proved the feasibility of applying transfer learning to identify the 4 Cerasus humilis cultivars in the real environment. In addition, considering that it will be deployed on the mobile terminal to identify the Cerasus humilis cultivars in the future, the lightweight model will be selected for research later.



a) Loss function value



b) Average accuracy

Fig. 4 - The accuracy and loss function on the verification set

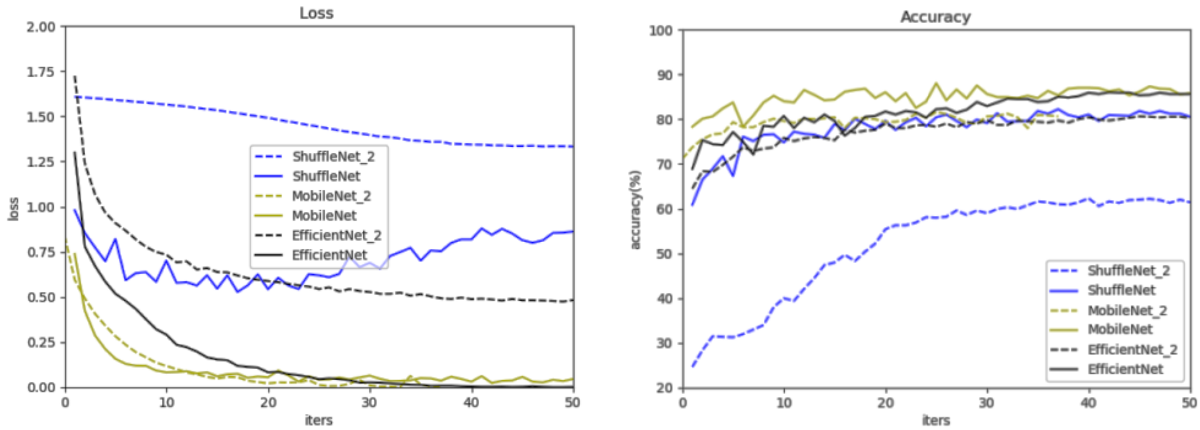
Fig. 4 shows the change of precision and loss function of each model in the validation set when the number of training iters was set to 50. From Fig. 4(a), it can be seen that the lightweight model EfficientNet-

B0 has a very smooth loss on the validation set, while the models MobileNetV3 and ShuffleNet have a relatively smooth loss on the validation set.

As shown in Fig. 4(b), the models showed that different characteristics in the validation accuracy change during the training. The overall recognition accuracy of EfficientNet-B0 and MobileNetV3 models was higher than ShuffleNet. MobileNetV3 had a higher verification accuracy in the initial iteration, and its accuracy shook during the iteration process. For EfficientNet-B0 with the number of iterations increases the accuracy had been showing an upward trend, and the shaking phenomenon was not obvious, indicating that the EfficientNet-B0 model, which had been co-expanded in depth, width and resolution, compared with the MobileNetV3 and ShuffleNet model has higher stability and can steadily improve the accuracy of the model.

The Second Transfer Learning Strategy

In this section will be given the testing results of 3 kinds of lightweight deep learning models by using the transfer learning strategy two, utilizing the pre-trained neural network architectures such as MobileNetV3, ShuffleNet, and EfficientNet-B0 as feature extractors, and the top layers are truncated by defining a new fully-connected layer with the practical number of classification and fine-tuned the higher dimensional layers.



a) Loss function value *b) Average accuracy*

Fig. 5 - The results of different transfer learning strategies

Fig. 5 showed the average precision and loss function values of different network models on the validation set when two transfer learning methods are applied. In Fig. 5, the first transfer learning method was represented by a solid line, and the second transfer learning method was represented by a dotted line. From the training results, it can be seen that the accuracy curves that for all models, fine-tuning the parameters of a pre-trained neural network architecture achieved higher classification accuracy as compared to using the neural network architecture with feature extraction only. Therefore, it is necessary to fine tune all layers of the network model using the transfer learning method.

Second, since the early stopping technique was used during the training, Fig. 5 showed that MobileNetV3 triggers early stopping in iteration 38 when using the neural network architecture as feature extractors only, while ShuffleNet and EfficientNet-B0 do not trigger early stopping during training. So the ShuffleNet and EfficientNet-B0 had better stability than MobileNetV3. From Fig. 5(a), it can be seen that the lightweight model EfficientNet-B0 has a very smooth loss on the validation set, while the model MobileNetV3 and ShuffleNet have a relatively smooth loss on the validation set, but the ShuffleNet has Serious oscillation.

The analysis results indicate that compared with other models, the EfficientNet-B0 model has the best classification accuracy in the four types of leaf datasets.

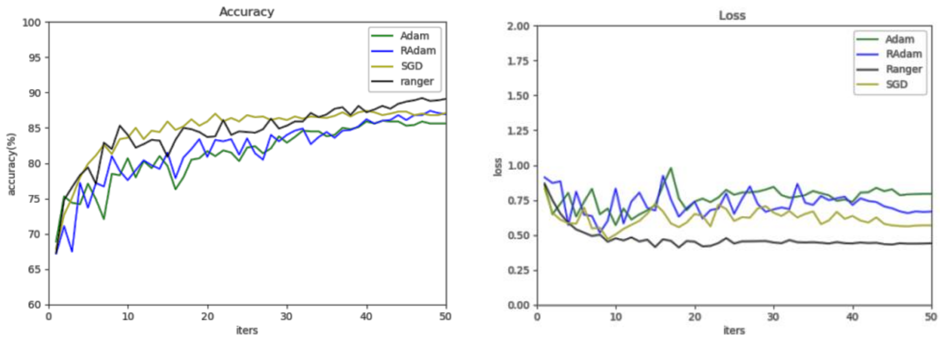
The Improved EfficientNet-B0

In this section, EfficientNet-B0 is used with Adam, RAdam, SGD, and Ranger optimizers respectively, and the accuracy of the models with different optimizers in the validation set and test set were shown in Table 3. In this section, the learning rate was set to 0.0001. The results in Table 3 showed that the accuracy of EfficientNet-B0 + Ranger was the highest in both the validation set and test set, and the other three had little difference.

Table 3

Comparison of EfficientNet-B0 with different optimizers

Model	Verification accuracy (%)	Test accuracy (%)
EfficientNet-B0+Adam	86.0	84.7
EfficientNet-B0+RAdam	87.4	85.3
EfficientNet-B0+SGD	87.4	85.1
EfficientNet-B0+Ranger	89.2	86.8



a) Accuracy rate curve *b) Loss curve*

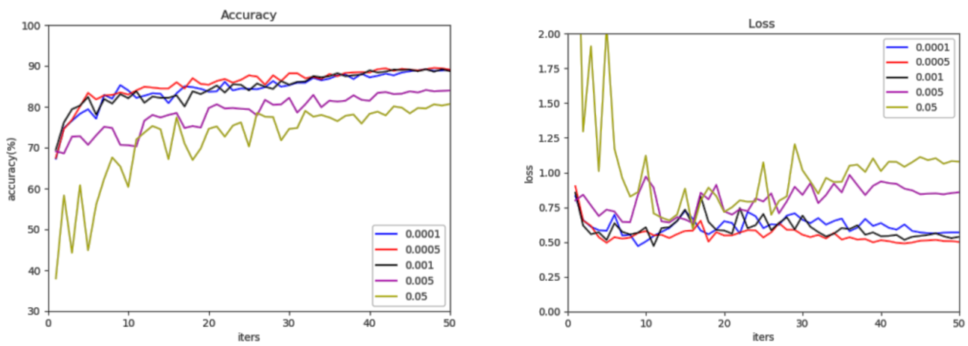
Fig. 6 - The accuracy and loss value of the model with different optimizers

Fig. 6 showed the changes in the accuracy of the EffiientNet-B0 with different optimizers in the verification set during 50 iterations. It can be concluded from Fig. 6 that the performance of the proposed EffiientNet-B0+Ranger outperformed all other pre-trained models achieving high overall accuracy of 86.8%. The reason was that EfficientNet-B0+Ranger with the optimizer of Ranger, which combines two very new optimizers (RAdam + Lookahead) into one optimizer, and had the advantages of RAdam and LookAhead at the same time. It can speed up the convergence speed and improve recognition accuracy. EfficientNet-B0+Ranger showed a steady increase in accuracy compared to other optimizers.

Effect of Learning Rate on The Model

Choosing the optimal learning rate is important as it determines whether the neural network can converge to the global minimum. It is used to control the update amplitude of weights and bias terms in order to achieve the optimal solution. The proper learning rate can accelerate the convergence of model training and achieve good classification results.

In order to find the most suitable learning rate for the Cerasus humilis cultivars leaf data set, five groups of learning rates (0.0001, 0.0005, 0.001, 0.005, and 0.05) were selected respectively for the EfficientNet-B0+Ranger model. Comparing the effects of different learning rates on the performance of the EfficientNet-B0+Ranger model, the comparison of model performance is shown in Fig. 7 (a) and (b).



a) Accuracy rate curve *b) Loss curve*

Fig. 7 - EfficientNet of learning rate on training accuracy and training loss

It can be seen from Fig. 7 that the convergence rate of the model does not accelerate with the increase of the learning rate. When the learning rate was 0.05 or 0.005, the curve oscillated seriously, the convergence was slow, and when the maximum number of iterations was reached, the model had not yet reached a convergence state. This indicates that when the learning rate was too high, it will cause the parameters to be optimized to fluctuate around the minimum value and not converge. When using a learning rate of 0.001, the loss value curve and average accuracy change curve of the model were more stable than the model curve when using a learning rate of 0.001 and 0.0001, and the model had a lower loss value and higher accuracy.

When fine-tuning the model, the performance of the model was better when the initial learning rate was set to a small value. The reason was that: in the transfer learning mode, all the layers of the network had been well trained, and the weight parameters of the model were close to the optimal solution, if the higher learning rate was used in the fine-tuning training phase, the model will skip the optimal solution and produce more oscillation, which will made the loss larger and the accuracy lower.

The network model of EfficientNet-B0+Ranger converged fastest and had higher accuracy when the learning rate was 0.0005, so the following research will choose the learning rate of 0.0005.

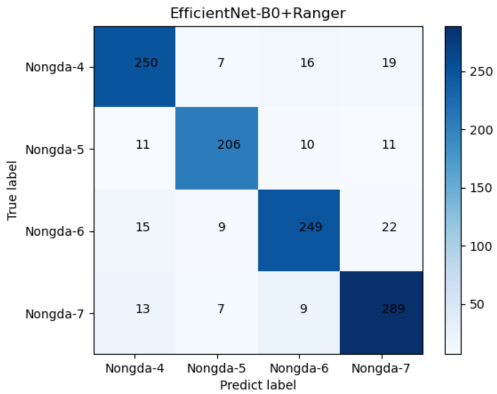


Fig. 8 - The confusion matrix of EfficientNet-B0+Ranger on the test set

To further evaluate the performance of the improved model, the confusion matrix of EfficientNet-B0+Ranger on the test set is shown in Fig. 8. The classification performance of the model was evaluated by precision, recall rate and average accuracy *(Powers et al., 2020)*, and the results were shown in Table 4. The Accuracy, Recall, and Precision, as defined in Equations (5-7).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$\mathrm{Re}\,call = \frac{TP}{TP + FN} \tag{6}$$

$$\mathrm{Pr}\,ecision = \frac{TP}{FP + TP} \tag{7}$$

Where true positive (TP) denotes the number of instances that the classifier accurately identifies, false negative (FN) is the reverse, which is the number of instances that are incorrectly identified, and false positive (FP) represents the number of cases that are mistakenly marked as such, and true negative (TN) is the number of samples that they are not in such disease and accurately identified by the classifier.

Table 4

The results of different metrics

Category name	Identified samples	Correct samples	Accuracy (%)	Precision (%)	Recall (%)
Nongda-4	292	250	85.6	95.0	86.5
Nongda-5	238	206	86.6	96.5	90.0
Nongda-6	295	249	84.4	94.6	87.7
Nongda-7	318	289	90.9	93.7	93.8

According to the confusion matrix of Fig. 8, the 4 Cerasus humilis cultivars leaves have the possibility of error, mainly because the 4 Cerasus humilis cultivars leaves are very similar, so if the image with low resolution or there is other interference factors will lead to more difficult to distinguish correctly.

As seen in Table 4, most of the samples in each class are correctly identified by the EfficientNet-B0+Ranger approach, and the average accuracy of the experimental results achieved 86.9% on the test set, which presents that the proposed approach had an impressive capability to recognize the Cerasus humilis cultivars in the field condition.

In a word, the experimental studies showed that EfficientNet-B0+Ranger models with the learning rate of 0.0005, offered the best classification performance and could be successfully used in Cerasus humilis cultivars recognition.

Discussion

Cultivar identification has long been dependent on visual recognition by experts, which highly relies on personnel skills. Therefore, accuracy cannot be guaranteed. It is of great significance to speed up the recognition process and make it accessible for non-experts. Although leaf image-based methods have been widely adopted in plant species identification. *Zhang et al. (2021)* proposed an Xception network to identify 88 sweet cherry cultivars and 100 soybean cultivars under controlled conditions, with an accuracy of 83.52% and 91.4%, respectively. *Liu et al. (2020)* proposed an efficient and convenient method for the classification of 14 apple cultivars with a sample background using a deep convolutional neural network and achieved an overall accuracy of 97.11%. However, the images they used were simple backgrounds, when transferring those methods from plant cultivars identification with simple backgrounds to the field of complex background, most of them performed worse, due to the high similarity of leaves among cultivars and other interfering factors in the field of complex background. As the field image acquisition is simple and has more practical significance, this study proposes an effective lightweight CNN for Cerasus humilis cultivars with complex background identification and achieved a high overall accuracy of 86.9%.

CONCLUSIONS

In this work, the lightweight EfficientNet-B0 deep learning architecture was proposed and compared with the state-of-the-art deep learning architectures ResNet50, GoogleNet, ShuffleNet, and MobileNetV3. The performance of the models with different transfer learning methods, different learning rates, and optimizers had been validated on an 1143 Cerasus humilis leaf dataset.

The experimental results confirmed that considering both the average accuracy and the model size on datasets, the lightweight models MobileNetV3, ShuffleNet, and EfficientNet-B0 were found to be superior to other traditional CNNs architectures. Furthermore, it was confirmed that for all models, fine-tuning the parameters of a pre-trained neural network architecture achieved better classification accuracy as compared to using the neural network architecture with feature extraction only. Moreover, the comparative experiments were carried out on EfficientNet-B0 with Adam, RAdam, and Ranger optimizers, and the experimental results confirmed that the optimizer of Ranger performed better with a faster convergence rate and higher accuracy compared to SGD, Adam, and RAdam. At last, the EfficientNet-B0+Ranger with the learning rate of 0.0005 had the best performance and achieved a high overall accuracy of 86.9%.

In summary, the improved EfficientNetB0+Ranger had smaller parameters and higher average accuracy, making it an ideal model for use in embedded devices.

In future work, it is planned to collect the hundreds of Cerasus humilis cultivars' leaf images from different planting areas to increase the generalization performance and efficiency of the model on more Cerasus humilis cultivars. By deploying these improved models in mobile environments, this presented model will also be used to identify other Cerasus humilis cultivars and even other plants. Furthermore, it is aimed to identify Cerasus humilis cultivars in real-time by studying other targets detection models such as Faster RCNN (Regions with Convolutional Neural Network), YOLO (You Only Look Once), and SSD (Single Shot MultiBox Detector).

ACKNOWLEDGEMENT

This work was supported by the Teaching and Research Management Fund Project of Shanxi Agricultural University, China (XK19207), Basic Research Project of Shanxi Province (Free Exploration) (Project No. 202203021212426).

REFERENCES

[1] Bai Z. H. (2015). Study on biological characteristics and drought resistance of cerasus humilis in the different site conditions and leaf size types. Inner Mongolia Agricultural University: Huhehot, China.

[2] Beikmohammadi, A., Faez, K., Motallebi, A. (2020). SWP-Leaf NET: a novel multistage approach for plant leaf identification based on deep learning. arXiv:2009.05139.

[3] Charters, J., Wang, Z., Chi, Z., Tsoi, A. C., Feng, D. D. (2014). Eagle: a novel descriptor for identifying plant species using leaf lamina vascular features. In 2014 IEEE international conference on multimedia and expo workshops (ICMEW), Chengdu, China, 14-18 July, 1-6.

[4] Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. Comput. Electron. Agric. 145, 311-318.

[5] Guo, X. Q., Fan, T. J., Shu, X. (2019). Tomato leaf diseases recognition based on improved multi-scale AlexNet. Transactions of the Chinese Society of Agricultural Engineering, 35(13), 162-169.

[6] John, S., Rose, A. L. (2021). Machine learning techniques in plant disease detection and classification-a state of the art. INMATEH-Agricultural Engineering, 65(3), 362-372.

[7] Kalyoncu, C., Toygar, Ö. (2015). Geometric leaf classification. Comput. Vis. Image Und. 133, 102-109.

[8] Li, J. Y., Zhao, Y. K., Xue, Z. E. (2019). A survey of model compression for deep neural networks. Chinese Journal of Engineering, 41(10), 1229-1239.

[9] Liu, C., Han, J., Chen, B., Mao, J., Xue, Z., Li, S. (2020). A novel identification method for apple (Malus Domestica Borkh.) cultivars based on a deep convolutional neural network with leaf image input. Symmetry, 12(2), 217.

[10] Liu, J., Yang, S., Cheng, Y., Song, Z. (2018). Plant leaf classification based on deep learning. In 2018 Chinese Automation Congress (CAC), Xi'an, China. 30 November-2 December, 3165-3169.

[11] Liu, Y., Feng, Q., Wang, S. Z. (2019). Plant disease identification method based on lightweight CNN and its mobile application. Transactions of the Chinese Society of Agricultural Engineering, 35, 194-204.

[12] Mouine, S., Yahiaoui, I., Verroust-Blondet, A. (2012). Advanced shape context for plant species identification using leaf image retrieval. In Proceedings of the 2nd ACM international conference on multimedia retrieval, New York, United States, 5-8 June, 1-8.

[13] Naresh, Y. G., Nagendraswamy, H. S. (2016). Classification of medicinal plants: An approach using modified LBP with symbolic representation. Neurocomputing, 173, 1789-1797.

[14] Powers, D. M. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv:2010.16061.

[15] Ramachandran, P., Zoph, B., Le, Q. V. (2017). Searching for activation functions. arXiv:1710.05941.

[16] Sun, J., Tan, W., Mao, H., Wu, X.; Chen, Y. (2017). Recognition of multiple plant leaf diseases based on improved convolutional neural network. Trans. Chin. Soc. Agric. Eng. 33(19), 209-215.

[17] Tan, J. W., Chang, S. W. (2018). Deep learning for plant species classification using leaf vein morphometric. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 17(1), 82-90.

[18] Zhang, P., Yang, L., Li, D. (2020). EfficientNet-B4-Ranger: A novel method for greenhouse cucumber disease recognition under natural complex environment. Comput. Electron. Agric., 176, 05652.

[19] Zhang, Y.; Peng, J.; Yuan, X.; Zhang, L. (2021). MFCIS: an automatic leaf-based identification pipeline for plant cultivars using deep learning and persistent homology. Hortic. Res., 8(1), 1-14.