Flower identification based on Deep Learning

Mengxiao Tian¹, Hong Chen¹, Qing Wang¹

¹China Agricultural University, 100083 No. 17 Qinghua East Road, China

norman_chen@263.net

Abstract. In the field of plant scientific research, agroforestry investigation and production and management, plant identification is crucial basic work, and flower identification is an important part of plant identification. Given the present artificial defects of labor cost, low efficiency and low accuracy in present artificial flower information query and traditional computer vision method, the study built a modified tiny darknet in flowers classification method. Seventeen types of flower datasets published by Oxford University are taken as the research objects and the input of the neural network model. The deep network classification model is trained to automatically extract the characteristics of flower images. Combined with softmax classifier, the flower test images are classified and identified. The experimental results show that the classification accuracy is 92% which is higher than the classification algorithm results of the original model and some current mainstream models. This model has a simple structure, few training parameters, and has achieved a good recognition effect. It is suitable for automatic classification and recognition in the field of flower planting and is convenient for the retrieval of agricultural plant information database.

1. Introduction

The identification of flowers plays a crucial role in agriculture, forestry and other related industries. In ecological agriculture, automatic identification of flowers is used to replace manual query of flower retrieval information database to improve the efficiency of retrieving flower information and reduce the loss of labor cost. In the scientific research, the forestry investigation, it provides efficiency support. For elementary and middle school students’ popular science education, horticulture appreciation, flower fair exhibition and so on, people wish to obtain the type of certain flowers and information in time, so it is significant in this aspect. The manual query is easy to make mistakes, and have a heavy workload, labor cost, material and financial resources. Researchers of traditional computer vision methods rely on the experience of the characteristics of artificial selection, and the design characteristics of the model generalization ability are not high enough. Due to the great differences among similar flowers, there are also great similarities among different types of flowers, such as the diversity of forms and changes in lighting conditions, uncomfortable factors of the background environment which cannot be solved well by the traditional computer vision method at present. However, deep learning can solve the problem of recognition in a complex environment without considering light change, complex feature extraction and specific environment.

At present, in the field of computer vision, the research achievements of flower classification mainly include Nilsback and Zisserman[1] et al., who put forward the method of automatic flower classification in many species, and designed four different features to be applied to flower image classification through manual feature extraction. Fernando[2] proposed a method of distinguishing feature fusion based on image classification, and in this method, the colour and shape features are fused by logistic regression
method for flower image classification. Gehler and Nowozin[3] et al., They assigned different weights to multiple image features to classify multiple objects. This design method can better identify flowers, but the design of the complex skills, steps are required. It relies on the experience of researchers to artificial selection. Moreover, the feature generalization ability of the designed model is not high due to the diversity of target morphology, the change of lighting conditions, the uncontrollable factors of the background environment, etc., which will have an impact on target recognition. These traditional computer vision algorithms have limitations and blindness. Due to the rapid rise of deep learning in recent years, it has gradually replaced traditional methods and been applied in various fields with very effective effects. The concept of deep learning proposed by Hinton[4] et al., in 2017, is a method of feature learning based on data. The essence of deep learning is to build a deep learning network model with many hidden layers, and a large number of training data to learn more useful features with fewer parameters and deeper structure. Compared with traditional machine learning, deep learning can automatically extract features, attach importance to the depth of model structure, and has excellent feature learning ability, thus making classification or prediction easier and more accurate. Currently, neural network models widely applied in the field of deep learning, such as AlexNet[5], VGG[6], DRN[7], GoodLeNet[8], SqueezeNet[9] and so on, have achieved good results in speech processing, natural language processing, image processing and other complex tasks, with high efficiency, plasticity and universality. Therefore, based on this background study, deep learning is applied to flower recognition and can identify flowers quickly and accurately. It can automatically learn image features instead of manually extracting features. Compared with traditional algorithms, it can improve the accuracy and stronger robustness.

Aiming at the limitations of the research methods and the complexity of image data, the study chose the flower classification model based on the tiny darknet[10-11] that was improved and whose structure parameters were slightly reduced, as shown in Fig.1. Seventeen types of flower data sets published by Oxford University were taken as the research objects. The flowers were quite common in the UK. These images had the characteristics of different sizes, shapes and illumination differences. Besides, there were great differences among these similar flowers and great similarities among different types of flowers. The flower data set is shown in Fig.2. This data set collected 17 different types of flowers images, with 80 images in each category and a total of 1360 images. Then, the average pixel filling method of the image was used to save the flower image with a size of 224×224 and a resolution of 96×96. In order to improve the quality of sample training and prevent overfitting, the transformation method of data enhancement is adopted to generate

![Fig.1 Flowers classification framework](image-url)

2. Materials and Methods

2.1. Flower image data set

In this study, Seventeen types of flower data sets published by Oxford University were taken as the research objects. The flowers were quite common in the UK. These images had the characteristics of different sizes, shapes and illumination differences. Besides, there were great differences among these similar flowers and great similarities among different types of flowers. The flower data set is shown in Fig.2. This data set collected 17 different types of flowers images, with 80 images in each category and a total of 1360 images. Then, the average pixel filling method of the image was used to save the flower image with a size of 224×224 and a resolution of 96×96. In order to improve the quality of sample training and prevent overfitting, the transformation method of data enhancement is adopted to generate...
more images to expand the image training set, and multiple combination transformations in data enhancement were sued to expand the data set and make the model more robust. In this study, the image is rotated to the left and right angels, and the orientation of the image content is changed. Secondly, each image along the horizontal direction of the left and right symmetry transformation as can be shown in Fig. 2. This algorithm increased each image of each category by 8, and each category increased form 80 to 640, and the data set increased from 1360 to a total of 10880. The flower image set was fixed into 2 data sets, namely the training set (10,200) and the test set (680) which were not intersected and maintained certain independence. The small training set and the large test set make the classification more difficult, and the data cannot be well fitted, which is prone to overfitting.

In Fig.2, first, the original image is rotated from 0 to 90 degrees, 90 degrees, 180 degrees, 270 degrees, etc., to obtain images (1), (2), (3), (4), and then these images are subjected to left-right symmetric transformation to obtain corresponding images (5), (6), (7), and (8). The enhanced data set is randomly divided into a training set and a test set, as shown in Table 1.

| Dataset          | Training set | Test set | Total  |
|------------------|--------------|----------|--------|
| Before augmentation | 1275         | 85       | 1360   |
| After augmentation | 10200        | 680      | 10880  |

2.2. Convolutional neural network
Convolutional neural network [12], is a multi-layer neural network deep learning method proposed by Yann LeCun form New York University in 1998. It is composed of multiple convolutional layers, and each layer is composed of multiple independent neurons. The basic structure of the neural network includes the convolutional layer and the down-sampling layer. First, in the convolutional layer, the neurons in each layer are locally connected with a neural unit in a small field of the previous layer, and the local features are extracted. Secondly, in the lower sampling layer, the feature graph is obtained form the convolutional layer for down-sampling, namely maximum pooling, and realizes the translation invariance of extracted features, simplifies the sample training parameters and reduces the complexity of network parameter selection. Compared with traditional methods, the advantage of CNN is that it can
automatically extract data features and automatically learn the high-level feature expression required by classification. In addition, weight sharing and local perception reduce the complexity of its network model. After training, features can be obtained with higher accuracy and faster speed.

2.3. Tiny darknet structure and improvements

The study selected tiny darknet classification convolution neural network as the network architecture. It is a simplified version of Darknet neural network, and the network structure is by Alexnet extensions of a lightweight architecture of CNN. Compared with LeNet-5[12], Alexnet, SqueezeNet and other convolution network model, the network model design, the network model have compact design and powerful functions with only 800 million floating point calculations, while other network is billions of floating point calculation. So it achieved more efficient distributed training. Tiny darknet is composed of multiple convolution layers which are similar to the multi-layer 3×3 convolution kernel of VGG model. A layer of 1×1 convolution kernel is inserted in the middle for dimensionality reduction, and the number of characteristic graphs of each layer is reduced, and the number of channels output from the pooling layer connected with the next layer is doubled. The number of parameters and calculation amount of the entire network model is less, which reduces the running time and model memory, thus shortening the training time of the model. The classification accuracy is also improved, with strong self-learning ability. Therefore, we choose this model as the basic model and make improvements the structure.

As shown in Fig.3, the modified version of tiny darknet consists of 15 convolutional layers and two BN layers (Batch Normalization) alternately. The input image is 224 × 224 × 3 (RGB three-channel image).

(1) Convolution layer

The first layer is a convolution layer with a depth of 16, each of which is connected to a small 3×3 domain in the 224×224 image. The convolution kernel is 3×3 with stride 2, which replaces all the original 2×2 max pooling layer for dimensionality reduction. This method[15] has been shown to work well in the cifar dataset, and could learn some of the necessary invariant features of the data. This convolution kernel learns image features as detectors and applies them anywhere in the image. "×2" means that this layer of network is repeated twice and connected to the next layer of network. The stride size is uniformly set to 1, and the output dimension of the entire network structure remains unchanged from the original model. Each 3×3 image patch is convolved in the 224×224 image, traversing each 3×3 neuron, the boundary padding is 1, and the convolved feature is secondarily activated by nonlinearity.
The way the function is mapped as the eigenvalue of a neuron in the convolutional layer. The activation function facilitates gradient training of deep neural networks, retains some image features, removes redundant data, and the number of feature maps output by the entire network architecture is consistent with the original model. The entire structure is matched with the rectified linear unit ReLU[16] function to remove redundant data, simplify the model parameters, and the convolution operation plus the bias terms, the convolution operation formula is as follows:

\[ y = \text{ReLU}(wx + b) \]  

(1)

Where \( x \) represents the input characteristics of a picture; \( w \) is the weight of the convolution kernel; the bias term is \( b \); \( y \) represents convolution results.

(2) BN layer

In the process of training the model, it is necessary to manually adjust the learning rate, parameter initialization, weight decay coefficient, requires a certain skill and much time, and the BN layer can be used to slowly adjust the parameters without deliberately. In this model, the convolutional layer is followed by the BN layer, which performs a certain linear transformation on the input of each layer. The normalized value distribution remains unchanged, automatically weakening the unimportant feature variables, and automatically from many feature variables. Extracting important feature variables, reduces the order of magnitude of feature variables, effectively reduces the dispersion of certain gradients and reduce the complexity of the model, thereby improving the robustness of the model and preventing the over-fitting effect.

(3) Output layer

Finally, the output layer has 1000 neurons, each corresponding to each category. Since there are only 17 species of flowers, the number of neurons is changed to 17, and this layer is connected to the avg pooling layer, and the eigenvalues are averaged. Since the full connection layer is easy to cause over-fitting due to the complicated number of parameters, the avg pool is used instead of the full connection layer, and the previous layer of the avg pooling layer has a depth of 1000 convolution layer, which reduces the calculation of parameters and improves the classification. Efficiency does not affect the performance of the classification, and its performance is consistent with the original tiny darknet.

(4) Loss function

Since there might be a gap between the valuation and the actual valuation of model, a loss function should be added to evaluate the gap between the valuation and the real value and try to minimize the loss. The smaller the loss is, the closer the model is to the result of the label. In this paper, cross entropy is adopted as the loss function, and the formula is as follows:

\[ L = - \sum_{\text{samples}} \sum_{\text{categories}} t_{ki} \log(y_{ki}) \]  

(2)

Where \( t_{ki} \) represents a sample; \( k \) belongs to the probability of category \( i \); and \( y_{ki} \) is the probability of predicted category \( i \) of sample \( k \) by the model; and this probability represents the value calculated by softmax[17].

For the multi-classification problem of this model, softmax loss function is the extension of logistic regression model on multi-classification problem. Multiple neuron outputs are mapped to the interval of (0-1), and multiple different values can be taken. The cumulative sum of these values is 1. The maximum value is selected as the prediction result of the model. If there are \( m \) labeled training samples, for the training set \( \{x^{(1)}, y^{(1)} \ldots, x^{(m)}, y^{(m)}\} \), its input characteristic is \( x^{(i)} \in R \), class mark is \( y^{(i)} \in \{0,1,\ldots,k\} \), for the test set \( x \), if the function is to estimate the probability value of each category, and value is \( p(y = j | x) \), the specific hypothesis function is as follows:
\[ p(y^{(i)} = j | x^{(i)}; \theta) = h_{\theta}(x^{(i)}) \]

\[
\begin{align*}
    p(y^{(i)} = 1 | x^{(i)}; \theta) &= \frac{e^{\theta_1 x^{(i)}}}{\sum_{j=1}^{k} e^{\theta_j x^{(i)}}} \\
    p(y^{(i)} = 2 | x^{(i)}; \theta) &= \frac{e^{\theta_2 x^{(i)}}}{\sum_{j=1}^{k} e^{\theta_j x^{(i)}}} \\
    &\vdots \\
    p(y^{(i)} = k | x^{(i)}; \theta) &= \frac{e^{\theta_k x^{(i)}}}{\sum_{j=1}^{k} e^{\theta_j x^{(i)}}}
\end{align*}
\]  

(3)

\[ \theta_1, \theta_2, \ldots, \theta_k \in \mathbb{R}^{n+1} \] is the model parameter, \[ \frac{1}{\sum_{j=1}^{k} e^{\theta_j x^{(i)}}} \] is to carry on data normalization to probability distribution, and make the sum of its probabilities equal to 1.

(5) Gradient descent algorithm

In this paper, the adaptive moment estimation method proposed by Adam[18] was adopted. Gradient descent method is mainly used to update the weight in the neural network model that means to update and adjust the parameters in one direction continuously to minimize the loss function. The Adam algorithm can calculate different adaptive learning rates for different parameters. It only keeps the exponential decay average of gradient square, but also keeps the decay value of previous gradient. Compared with other gradient descent algorithm, it has faster convergence speed, small memory requirement and better adaptive effect. In addition, problems such as disappearance of learning rate, slow convergence and large fluctuation of loss function performance when parameters are updated can be corrected. It will not cause the step size to become larger due to a large gradient, and the value of parameter is relatively stable, which make the trained learning model more loaded with image features.

The effects of the Adam gradient descent algorithm and the SGD[19] stochastic gradient descent algorithm after running 35001 steps in the model are shown in Fig. 4 and Fig. 5.

![Adam gradient descent algorithm](image1)

![SGD stochastic gradient descent algorithm](image2)

Fig.4 Adam gradient descent algorithm            Fig.5 SGD stochastic gradient descent algorithm

Note: Accuracy indicates the accuracy of the test set, and Test loss represents the loss function of the test set.

From the convergence situation of test loss in Fig.5 and Fig.6, it can be seen that Adam is faster than SGD in convergence, and the convergence effect is relatively stable. It can correct the initial deviation, which is better than SGD, so the Adam gradient is used. The algorithm trains the learning model.

3. Experiment set up

This paper builds a neural network model based on Google's second-generation artificial intelligence learning system TensorFlow [20] framework, which functionalizes each algorithm, directly invokes it and then feeds it to the artificial intelligence neural network for analysis and processing. Widely used in
product development and scientific research in various fields. This experiment prepares 10200 training sets and 680 test sets, where the training set and the test set are the same data source. There is no crossover between data sets and maintain a certain degree of independence. At the beginning of the training, we randomly weight the image features and take a Gaussian distribution to initialize the weight matrix $w$. The bias term is a standard normal distribution of 0.0005 (mean value is 0, variance is 1), and the initial learning rate of the weight is set to 0.0001. In order to save training time and make the model converge faster, we use the Adam gradient descent algorithm to train the optimization model. This training converges in about 2 hours. The NVIDIA Tesla k40m graphics card is used as the computing platform, and the operating system is windows 64-bit.

4. Experimental results and analysis
In order to test the effectiveness of the proposed method, the test set is obtained by the trained deep learning model and using the softmax classifier for classification. We can obtain the sample classification recognition rate of 92%. The accuracy rate of each category is shown in Table 2.

| Flower types     | Number | Recognition accuracy |
|------------------|--------|----------------------|
| Buttercup        | 42     | 86.50%               |
| Colts Foot       | 39     | 87.80%               |
| Daffodil         | 41     | 76.85%               |
| Daisy            | 33     | 100%                 |
| Dandelion        | 38     | 89.60%               |
| Fritillary       | 40     | 94.80%               |
| Iris             | 44     | 96.50%               |
| Pansy            | 39     | 91.50%               |
| Sunflower        | 34     | 100%                 |
| Windflower       | 41     | 85.40%               |
| Snowdrop         | 44     | 88.50%               |
| Lily valley      | 35     | 94.50%               |
| Bluebell         | 44     | 95.50%               |
| Crocus           | 46     | 97.85%               |
| Tigerlily        | 45     | 97.80%               |
| Tulip            | 42     | 77.50%               |
| Cowslip          | 33     | 95.50%               |

From the data in Table 2, the identification rate of classification with flowers is higher in this paper. There are 17 kinds of flowers in the image, totalling 680 images. Among them, the classification recognition rate of daisy and sunflower has reached 100%, and there are 15 categories of 80%. It shows that our model has good classification and recognition results.

As can be seen from Table 2, only tulip and daffodil do not reach 80% recognition rate, and other flower accuracy is as high as 80% or more. Therefore, eight representative images are selected from the test set for analysis, as shown in Fig.6. GT is the true value of the label of the flower, and Pred is the predicted classification result. The first column is to predict the correct flower. The first column is to predict the correct flower, from which the illumination changes (the second picture in the first column), the partial occlusion (the first picture in the first column), and the background change, environmental noise (the third and fourth pictures in the first column), the classification has a good effect and robustness. The first two image truth labels in the second column are daffodil. The predicted results are cowslip and tulip. Especially the second image of the second column. In the third image, the tulip and daffodil flowers have high similarity.
It can be seen that there is still a certain similarity between these images, and the interference of the background environment is likely to lead to prediction errors, the second picture in the fourth line predicts a complete error. There is a big difference between these two types of flowers. The model designed in this paper does not have a good generalization. So our next step is to improve the model with the Generative Adversarial Networks, and extend the sample data of similar data sets, further improve the recognition rate of flowers.

Table 3 This method is compared with the present method

| Research method | Accuracy |
|-----------------|----------|
| Literature[1]   | 88.3%    |
| Literature[2]   | 91.0%    |
| Literature[21]  | 86.93%   |
| Literature[22]  | 80.0%    |
| Xception        | 93.38%   |
| Inception-v3    | 96.81%   |
| ResNet50        | 59.80%   |
| VGG16           | 88.24%   |
| VGG19           | 88.73%   |
| Tiny darknet    | 90.5%    |
| Our model       | 92%      |

In order to test the classification performance of the deep learning network model, we compare with the mainstream research methods in the academic world, using the Oxford 17flower data as an experimental sample. The comparison of the experimental results is shown in Table 3: It shows that the classification method of the deep learning model in this paper is better than the original tiny darknet model, indicating that our improved model improves the generalization ability of the network. The stickiness and small error are lower than the recognition rate of Xception and Inception-v3 network models, which is higher than that of other network models, and has better classification and recognition.
results. The recognition rate of the literature[2] is close to the recognition rate of the method in this paper, but the feature extraction of the image of the flower is manual feature extraction, and the designed generalization ability is not high and the method steps are complicated. The method of this paper automatically selects the feature of the flower image, without paying attention to the preprocessing of the image data, directly normalizing the image to the same size, and then inputting the deep learning model, the operation is simple, and the flower feature can be well recognized, and has practicality; the X-ception model and the Inception-v3 model are higher than the model recognition rate, but the network model design is not as small as the tiny darknet model adopted in this paper. The parameter size is large, and the model memory is dozens of times of the tiny darknet model. It is not able to optimize the parameter counting well, and the model storage is small and the portable mobile end has mobility.

5. Conclusion
At present, the flower classification method in the field of computer vision has complicated operation and low classification accuracy. In this paper, a recognition method based on in-depth learning is proposed. Its application significances and advantages are as below:

(1) The biggest advantage of this model of tiny darknet modified version is fast, light-weight network, small model, optimized parameter calculation, and more accurate classification identification.

(2) It can be directly applied to the agricultural flower planting industry for classification and identification, to facilitate the retrieval of agricultural plant information database.

(3) For the popular science education and garden appreciation of primary and secondary school students, it is possible to quickly obtain which type of flower and related attribute information of a certain flower.

Our experiments show that the method is efficient and the recognition rate is high. The next step is to research the network to solve the problem of insufficient flower datasets, and to add Chinese flower datasets to identify more complex flowers, and to recognize very similar flowers, combined with migration learning on mobile. The application is deployed for promotion.

Acknowledgments
This work was supported by National Natural Science Foundation of China [Grant No.41601491];The Fundamental Research Funds for the Central Universities [Grant No. 2018QC080, 2018TY002]).

References
[1] Nilsback, Maria Elena, and A. Zisserman. "Automated Flower Classification over a Large Number of Classes." Sixth Indian Conference on Computer Vision, Graphics & Image Processing IEEE Computer Society, 2008:722-729.
[2] Sebban, M., Muselet, D., Fromont, E., & Fernando, B. (2012). Discriminative feature fusion for image classification. Computer Vision and Pattern Recognition (Vol.157, pp.3434-3441). IEEE.
[3] Gehler, Peter, and S. Nowozin. "On feature combination for multiclass object classification." IEEE, International Conference on Computer VisionIEEE, 2010:221-228.
[4] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786):504.
[5] Krizhevsky, Alex, I. Sutskever, and G. E. Hinton. "ImageNet classification with deep convolutional neural networks." International Conference on Neural Information Processing Systems Curran Associates Inc. 2012:1097-1105.
[6] Simonyan, Karen, and A. Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition." Computer Science (2014).
[7] He, Kaiming, et al., "Deep Residual Learning for Image Recognition." (2015):770-778.
[8] Szegedy, Christian, et al., "Going deeper with convolutions." Computer Vision and Pattern Recognition IEEE, 2015:1-9.
[9] Iandola F N , Han S , Moskewicz M W , et al., SqueezeNet: AlexNet-level accuracy with 50x fewer
parameters and <0.5MB model size[J]. 2016.
[10] Redmon J, Divvala S, Girshick R, et al., You Only Look Once: Unified, Real-Time Object Detection[J]. 2015.
[11] Redmon, Joseph, and A. Farhadi. "YOLOv3: An Incremental Improvement." (2018).
[12] Lecun Y L, Bottou L, Bengio Y, et al., Gradient-Based Learning Applied to Document Recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.
[13] Simon M, Rodner E, Denzler J. ImageNet pre-trained models with batch normalization[J]. 2016.
[14] Ma Y, Klabjan D. Convergence Analysis of Batch Normalization for Deep Neural Nets[J]. 2017
[15] Springenberg J T, Dosovitskiy A, Brox T, et al., Striving for simplicity: The all convolutional net[J]. arXiv preprint arXiv:1412.6806, 2014.
[16] Glorot, Xavier, A. Bordes, and Y. Bengio. "Deep Sparse Rectifier Neural Networks." Jmlr W & Cp 15(2011).
[17] Liu, Weiyang, et al., "Large-Margin Softmax Loss for Convolutional Neural Networks."//proceeding of the 33rd International Conference on Machine Learning, 2016:507-516.
[18] Kingma, Diederik P, and J. Ba. "Adam: A Method for Stochastic Optimization."Computer Science(2014).
[19] Bottou, Léon. Large-Scale Machine Learning with Stochastic Gradient Descent. Proceedings of COMPSTAT'2010, Physica-Verlag HD, 2010:177-186.
[20] TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems[J]. 2016.
[21] ZHOU W,WU G S. Research on saliency map based flowerimageclassification algorithm[ J ]. Computer Technology and Development.
[22] Xiao-Xin W U, Liang G, Min Y, et al., Flower species recognition based on fusion of multiple features[J]. Journal of Beijing Forestry University, 2017.