Flower image classification based on generative adversarial network and transfer learning

Xiaoxue Li¹, Rongxin Lv², Yanzhen Yin³, Kangkang Xin⁴, Zeyuan Liu⁴ and Zhongzhi Li⁵*

¹ College of Electronic and Information Engineering, Shenyang Aerospace University, Shenyang, Liaoning, 110136, China
² College of Innovation and Entrepreneurship, Shenyang Aerospace University, Shenyang, Liaoning, 110136, China
³ College of Science, Shenyang Aerospace University, Shenyang, Liaoning, 110136, China
⁴ Academy of Aeronautics and Astronautics, Shenyang Aerospace University, Shenyang, Liaoning, 110136, China
⁵ College of Computer Science, Shenyang Aerospace University, Shenyang, Liaoning, 110136, China

*Corresponding author’s e-mail: sau_lzz@email.sau.edu.cn

Abstract. Aiming at the problem that the classification accuracy of the traditional flower classification method is low and the deep neural network requires a large amount of original data. This paper designs a flower classification model that combines generative adversarial network and ResNet-101 transfer learning algorithm, and uses stochastic gradient descent algorithm to optimize the training process of the model. The experimental results on the international public flower recognition dataset, Oxford flower-102 dataset, show that by enhancing the original data, the accuracy of the network's recognition and classification of flowers is improved. At the same time, the model proposed in this paper is superior to other traditional network models, with higher recognition accuracy and robustness.

1. Introduction

Aiming at the problem that the classification accuracy of the traditional flower classification method is low and the deep neural network requires a large amount of original data. This paper designs a flower classification model that combines generative adversarial network and ResNet-101 transfer learning algorithm, and uses stochastic gradient descent algorithm to optimize the training process of the model. The experimental results on the the international public flower recognition dataset, Oxford flower-102 dataset, show that by enhancing the original data, the accuracy of the network's recognition and classification of flowers is improved. At the same time, the model proposed in this paper is superior to other traditional network models, with higher recognition accuracy and robustness.

Currently, flower image classification methods can be divided into two categories: methods based on manual feature extraction and methods based on deep learning to automatically extract features. Manual feature extraction methods mostly extract color features, texture features, and shape features of images, and combine them with machine learning algorithms for classification. For example,
Nilsback et al. [4] created a new dataset containing 102 types of flowers, Oxford 102 Flowers, and they also proposed to use support vector machine (SVM) sort with combing the scale-invariant feature transform (SIFT) features and the orientation gradient histogram (HOG) features. The classification accuracy of this method on the Oxford 102 Flowers dataset is 72.8%. Literature [5] encodes the local information and spatial information of the feature context, and uses support vector machine for classification. Xie Xiaodong et al. [6] proposed a method that combines the salient features of flowers with the Grabcut algorithm, using multi-feature fusion, and finally applies Support Vector Machine (SVM) for classification. However, traditional methods based on manually extracting feature points are not sufficient to extract features, resulting in low classification accuracy. While the method of automatically extracting based on deep learning requires a higher amount of experimental data. And due to the large number of training parameters of the network, the neural network is prone to underfitting.

Generative adversarial network (GAN) is a generative model proposed in 2014 [7], which trains generators and discriminators based on random noise and real samples to imitate fake and real images, which can achieve data enhancement. On the basis of GAN, many variant frameworks and networks can be better applied in various directions. Among them, GAN-based structural variants such as Conditional Generative Adversarial Network (CGAN), Deep Convolutional Generative Adversarial Network (DCGAN), and Cyclic Generative Adversarial Network (CycleGAN) have been widely studied and applied in the image field. The mapping relationship between the input and output domains of the high-dimensional model of the generative adversarial neural network model shows excellent performance [8-10].

This paper proposes a flower recognition method based on generative adversarial network and transfer learning. Through the generation of adversarial network, the data of fake and real images are used to enhance that of flower images. ResNet-101 transfer learning algorithm is applied to the original data set and the amplified data set. The experimental results show that this method effectively improves the recognition accuracy of imbalanced datasets.

2. Model

2.1. The proposed model

In order to overcome the problem of the imbalance in the number of different samples in the sample dataset, this paper proposes to use a generative confrontation network to expand the original data set, and then combine the ResNet-101 migration learning model to classify and recognize flowers. The overall framework of the model proposed in this paper is shown in Figure 1.
2.2. Generative Adversarial Network

The Generative Adversarial Network (GAN) was proposed by Goodfellow et al., which consists of a pair of generator and discriminator. The basic idea is to find the Nash equilibrium point in the high-dimensional parameter space through the mutual game and confrontation between the generator and the discriminator. The idea of GAN network construction is shown in Figure 2. The generator \( G(x) \) represents the mapping relationship \( G(x) \) from the input image to the generated image. The discriminator \( D \) is used to identify whether the data comes from the model \( D(G(x)) \) generated by the generator or the practical model \( D(y) \). The advantage of GAN lies in the low requirements for data distribution assumptions, which can not only adapt to the mapping of high-dimensional parameters but also break through the limitations of Gaussian distribution. In the generative adversarial neural network, the cost function \( V(G,D) \) of generator \( G \) and discriminator \( D \) is expressed as

\[
\min_G \max_D V(D,G) = E_{x\sim P_d(x)} \times [\log D(x)] + E_{z\sim P_z(z)} [\log(1 - D(G(z)))]
\]

In the formula, \( E_{x\sim P_d(x)} \) and \( E_{z\sim P_z(z)} \) represent the expectation from the real data and the expectation from the generator data, \( x \) represents the data of real flower picture inputted, subjects to the distribution \( P_d(x) \), \( z \) is the random noise input to the generator network \( G \), subjects to the distribution \( P_z(z) \), \( G(z) \) is the generated picture of the generator network. The purpose of the \( G \) network is to make \( D(G(z)) \) as large as possible and \( V(D,G) \) to be the smallest. The purpose of the \( D \) network is to make \( D(x) \) as large as possible and \( D(G(z)) \) as small as possible which means making \( V(D,G) \) be the largest.

Figure 1. The proposed framework of flower classification network
2.3. ResNet structure

ResNet proposes a residual network structure, adopts batch normalization, and sets the activation function of ReLU, which increases the depth of the network while avoiding performance degradation caused by gradient disappearance or explosion. The improved network proposed in this paper obtains deep features with rich semantic information through ResNet, which can better classify flowers in images.

As the number of network layers increases, the overall model’s representation ability becomes stronger, but the training accuracy drops instead, and the problem of network degradation appears. Residual learning can be used to solve the general degradation problem of deep networks. The expression of the residual function is:

\[ F(x) - H(x) - x \] (2)

In the formula, \( x \) is the input of the network, and \( H(x) \) is the basic mapping of multiple nonlinear network layers.

\[ y_i = x_i + F(x_i, \omega_i) \] (3)

\[ x_{i+1} = g(y_i) \] (4)

In the formula, \( x_i, x_{i+1} \) and \( \omega_i \) respectively represent the input, output and weight of the \( i \)-th residual block, \( F \) represents the residual function, and \( g \) represents the ReLU activation function.

From the above formula, the process of deepening the number of network layers can be expressed as:

\[ x_L = x_i + \sum_{j=i}^{L-1} F(x_j, \omega_j) \] (5)

Perform back propagation through the loss function to update the parameters. If the loss function is represented by \( \text{Loss} \), then

\[ \frac{\partial \text{Loss}}{\partial x_i} = \frac{\partial \text{Loss}}{\partial x_L} \ast \frac{\partial x_L}{\partial x_i} = \frac{\partial \text{Loss}}{\partial x_L} (1 + \frac{\partial F(x)}{\partial x}) \] (6)

In summary, when \( F(x) = 0 \), the superimposed nonlinear network layer is constructed as an identity mapping, so the features learned through the network are similar to the network input. In the actual network training environment, \( F(x) \neq 0 \), the network layer learns new features instantly, and the performance is better.
The structure composed of the non-linear network layer leads to a direct correlation channel between the input and the output, so that the network parameters are concentrated in learning the residuals. The residual block can be expressed as:

\[ Y = g[F(x, \{ \omega_i \}) + x] \]  

(7)

In the formula, \( x \) is the network input and \( Y \) is the network output. \( F(x, \{ \omega_i \}) \) represents the learned residual mapping, and \( g \) represents the ReLU activation function.

3. Experiment and analysis

3.1. Experimental configuration and data set

The computer environment used in this experiment is Intel(R) Core(TM) i5-8300H CPU@2.30 GHz, GeForce GTX 1050Ti graphics card, 4GB video memory, 8GB memory, the programming environment is Python 3.6, and the deep learning tool used is Keras.

This experiment uses the Oxford flower-102 public data set, which comes from the flower image database created by the Oxford University Visual Geometry Group. It contains 102 categories of flowers, each category has between 40-258 pictures, a total of 8189 pictures. The database also takes into account all the difficulties in the field of image recognition, such as complex backgrounds, many flower types and complex color changes, and some different flowers are highly similar. Therefore, the research on flower image classification is of great significance. Part of the data set is shown in Figure 3.

3.2. Analysis of Generate image data

In order to verify the effect of generative adversarial network on flower image data enhancement, this paper uses Python for example simulation. The experimental results are shown in Figure 4, where Figure 4(a) shows the generated image results of the network when the number of model training iterations is 0; Figure 4(b) shows the model training iterations of 50 times, Figure 4(c) shows the generated image result of the generated network when the number of model training iterations is 200; Figure 5(d) shows the generated image result of the generated network when the number of iterations
is 500. It can be seen from Figure 5 that as the model training process proceeds, the effect of the generated image is getting closer and closer to the real image.

![Generated Images](image1.png)

**Figure 4. Results of generate images.**

### 3.3. Result of Comparative Experiment

In order to prove the validity of the model proposed in this paper, relevant verification experiments are carried out. The number of iterations in the model training is set to 50, and the change in loss value and the change in classification accuracy are shown in Figure 5. It can be observed in the Figure 6 that the model loss value decreases rapidly during the training process. When the number of training iterations of the model is about 25, the classification accuracy of flowers has reached a result of greater than 90%. The classification accuracy of different models is shown in Table 1. It can be observed in the table that the classification accuracy of the traditional model is low. At the same time, after data enhancement, the prediction accuracy of each model has been improved to a certain extent. This reflects the effectiveness of the data augmentation strategy adopted by the generative adversarial network.
Figure 5. The training and verification process of the proposed model.

Table 1. Target detection results of different models

| Model             | Test result 1 | Test result 2 | Test result 3 | Test result 4 |
|-------------------|---------------|---------------|---------------|---------------|
| Non-Data enhancement CNN | 77.6%         | 76.9%         | 77.8%         | 78.1%         |
| Improved Alex Network       | 84.5%         | 83.9%         | 85.4%         | 84.4%         |
| GAN-TL                  | 88.9%         | 87.9%         | 88.8%         | 89.1%         |
| Data enhancement CNN     | 79.5%         | 80.4%         | 78.9%         | 79.8%         |
| Improved Alex Network       | 85.8%         | 84.9%         | 86.6%         | 86.1%         |
| GAN-TL                  | 89.6%         | 90.2%         | 88.4%         | 90.7%         |

4. Conclusion

The complexity of the environment where flowers are located in flower images, the diversity of flower categories, and the serious imbalance of data between categories make the accuracy of flower classification unsatisfactory. Aiming at the problem of imbalanced flower image categories, this paper proposes a flower classification model based on generative adversarial network and ResNet-101 transfer learning model, and uses stochastic gradient descent algorithm to update network weights. Experiments on the Oxford Flower 102 dataset show that the model achieves a higher classification accuracy by balancing the number of flowers among different categories. At the same time, it has better classification results than traditional methods and other deep neural network architectures. Its training classification accuracy reaches 96.2%, which proves the accuracy and feasibility of this method in flower image classification tasks.

Acknowledgments

This research was supported by 2020 Shenyang Aerospace University Students' Innovative Training Program (Project Number: S202010143042).
References
[1] Saitoh T, Aoki K, Kaneko T. Automatic recognition of blooming flowers[C]/Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004. IEEE, 2004, 1: 27-30.
[2] Gogul I, Kumar V S. Flower species recognition system using convolution neural networks and transfer learning[C]/2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN). IEEE, 2017: 1-6.
[3] Siraj F, Salahuddin M A, Yusof S A M. Digital image classification for malaysian blooming flower[C]/2010 Second International Conference on Computational Intelligence, Modelling and Simulation. IEEE, 2010: 33-38.
[4] Nilsback M E, Zisserman A. Automated flower classification over a large number of classes[C]/2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing. IEEE, 2008: 722-729.
[5] Qi W, Liu X, Zhao J. Flower classification based on local and spatial visual cues[C]/2012 IEEE international conference on computer science and automation engineering (CSAE). IEEE, 2012, 3: 670-674.
[6] Xie Xiaodong. Research on fine image classification for flower images[D]. Xiamen: Xiamen University, 2014.
[7] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]/Advances in neural information processing systems. 2014: 2672-2680.
[8] Sun W, Durlofsky L J. A new data-space inversion procedure for efficient uncertainty quantification in subsurface flow problems[J]. Mathematical Geosciences, 2017, 49(6): 679-715.
[9] Zhu Y, Zabaras N. Bayesian deep convolutional encoder–decoder networks for surrogate modeling and uncertainty quantification[J]. Journal of Computational Physics, 2018, 366: 415-447.
[10] Mo S, Zhu Y, Zabaras N, et al. Deep convolutional encoder - decoder networks for uncertainty quantification of dynamic multiphase flow in heterogeneous media[J]. Water Resources Research, 2019, 55(1): 703-728.