Mathematical Biosciences and Engineering

MBE, 18(3): 1981–1991.
DOI: 10.3934/mbe.2021103
Received: 02 November 2020
Accepted: 04 February 2021
Published: 25 February 2021

http://www.aimspress.com/journal/MBE

Research article

Visual attentional-driven deep learning method for flower recognition

Shuai Cao[1] and Biao Song[2,*]

[1] School of Information Science & Engineering, Lanzhou University, Lanzhou 730000, China
[2] Nanjing University of Information Science and Technology, Nanjing 210044, China

* **Correspondence:** Email: caosh18@lzu.edu.cn.

Abstract: As a typical fine-grained image recognition task, flower category recognition is one of the most popular research topics in the field of computer vision and forestry informatization. Although the image recognition method based on Deep Convolutional Neural Network (DCNNs) has achieved acceptable performance on natural scene image, there are still shortcomings such as lack of training samples, intra-class similarity and low accuracy in flowers category recognition. In this paper, we study deep learning-based flowers' category recognition problem, and propose a novel attention-driven deep learning model to solve it. Specifically, since training the deep learning model usually requires massive training samples, we perform image augmentation for the training sample by using image rotation and cropping. The augmented images and the original image are merged as a training set. Then, inspired by the mechanism of human visual attention, we propose a visual attention-driven deep residual neural network, which is composed of multiple weighted visual attention learning blocks. Each visual attention learning block is composed by a residual connection and an attention connection to enhance the learning ability and discriminating ability of the whole network. Finally, the model is training in the fusion training set and recognize flowers in the testing set. We verify the performance of our new method on public Flowers 17 dataset and it achieves the recognition accuracy of 85.7%.

Keywords: deep learning; feature extraction; attention learning; flower recognition

1. Introduction

The main purpose of flower recognition is to make judgments of flower category though some flower attributes, such as color, texture and semantics, which plays an important role in the fields of forestry informatization and plant medicine [1]. Different from classical image recognition [2–4],

flower category recognition is a typical fine-grained image recognition task, which requires model have strong inter-class and intra class discrimination capabilities, and also is a popular research topic in the fields of computer vision, pattern recognition and forestry informatization.

In recent years, deep learning [5] has achieved great success in computer vision, multimedia signal processing and natural language processing [6]. Although there are many classification methods in the literature [7–9], Deep Convolutional network (DCCN), as the most outstanding representative of deep learning, has been widely used in image classification, scene recognition, semantic information extraction, and still maintains the current best results [10,11]. In the view of this, some researchers have applied DCNN to the problem of flower category recognition and achieved good performance [12,13]. Although the methods based on DCNN can improve the accuracy and speed of flower category recognition, it still has 3 main problems: 1) The number of training samples is insufficient. DCNN always contains a lot of parameters, and training deep models in a small dataset is much challenging due to the over-fitting problem. Unfortunately, there have no public dataset with sufficient types and quantities at the same time in flower category recognition task, which directly limits the performance of the model. Even if the problem can be mitigated by data augmentation or fine-tuning on ImageNet, the useful information contained in the dataset is not increased. 2) Low recognition accuracy. Flower category recognition is a fine-grained image recognition task and has the characteristics such as high similarity between heterogeneous flowers. In addition, due to the complexity and variability of the natural environment, the pose and view angle of flowers may change unpredictably, which makes model difficult to train and the performance poor. 3) The background of image is complicated. Flower images collected from nature always have complex backgrounds and contain many noise, which may limit the recognition performance of deep learning models.

To enable deep learning quickly focus on the key points of the input data, self-attention mechanism-based model has been developed and successfully applied to many tasks, such as natural language processing and human-machine dialogue [14]. Human can quickly scan the visual information and obtain the attention target area. Not only that, but people pay more attention to the target area to get more detailed information and suppress other useless things, which is a survival mechanism formed by humans over a long period of evolution. Therefore, we proposed a novel flower recognition method based on attention mechanism (Visual Attentional-driven DCNNs, VA-DCNNs), which can effectively identify flower species accurately. The model is mainly divided into four-fold. Firstly, due to deep learning method always need massive training data to guarantee the performance, we adopt data augmentation techniques to increase samples. We rotate the picture in a clockwise angle and clip along the middle, which will be fuse with original samples as the training set for the experiments. Secondly, a Visual Attentional Learning (VAL) block are constructed for the vanilla DCNNs (we use ResNet14 and ResNet50 as the baseline in this paper), which makes VA-DCNNs have strong discriminative learning ability. Thirdly, the layer weights of the model are obtained by using dataset training. And finally, we get the recognition accuracy of the model on testing set. The experimental on the Flowers 17 public dataset prove the effectiveness of VA-DCNNs, which can achieve an accuracy of up to 85.7%. Compared with other recognition methods, VA-DCNNs can achieve better results, and has strong practicability and generalization.

2. Data augmentation

2.1. Flowers 17 dataset

The experimental dataset is the Flowers 17 [15], which contains 17 common flowers in the UK. The flowers including sunflowers, hyacinths, daffodils and chrysanthemums, etc., and each category have 80 images with different pose, size and perspective. Figure 1 have shown some examples in Flowers 17 dataset. So far the dataset has been widely used in flower recognition and organ segmentation, which is one of the most representative dataset in this field.

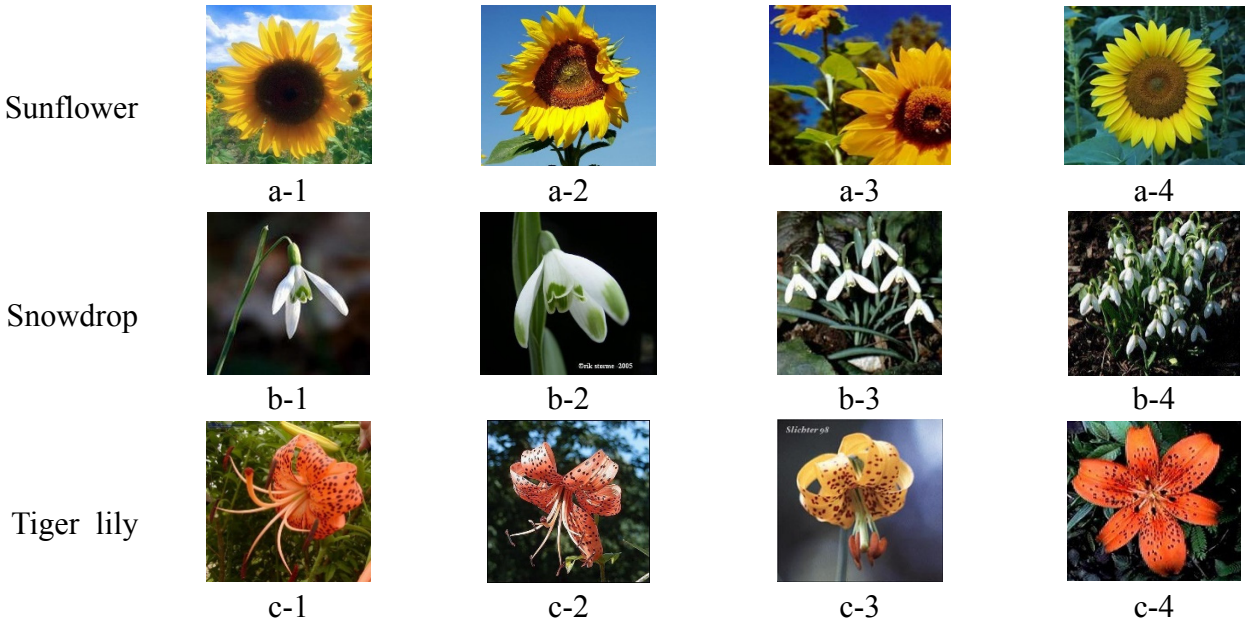


Figure 1. Some samples in Flowers 17 dataset.

2.2. Data augmentation

Deep convolutional neural network always composed by multiple blocks, and each block contains several convolutional layers, batch normalization layers, activation layers, and pooling layers. The data flows and gradient transfer between blocks through convolution kernels and back propagation algorithm. A DCNNs model always contains a large number of parameters need to be trained, which can make DCNNs fitting the data well. Although sufficient training samples can make the model fine-training, augmenting the training set to increase the number of training samples is one of the most common techniques used in deep learning models to further enhance the generality and robustness of the model [16]. In this paper, we augment the original flower image by using rotating and forward cropping. Specifically, for each original 224 × 224 pixel flower image, we rotate it clockwise, perform forward cropping every 30° and save it as 224 × 224 pixel size. The rotate operation totally perform 4 times (30°, 60°, 90°, 120°), and obtain a new dataset with five times than original set in quantity. The new dataset is divided into 70% training set, 20% validation set and 10% testing set, randomly. Since this data augmentation technique has been widely used in several papers, we don't repeat it here, but more details can be found in [11].

3. Block definition and model structure

3.1. Attentional-driven residual block

The traditional DCNNs module extracts features by using stacking convolutional, dropout, batch normalization and activation layers (as shown in Figure 2(a)). Although the effectiveness of this structure has been verified in many DCNNs models, single stack the block easily causes "gradient explosion" or "gradient disappearance" during training when network depth further increase. Deep layer blocks cannot take the input information or gradient is lost in the back-propagation process, resulting in the model cannot be trained [17]. Therefore, deep residual network (as shown in Figure 2(b)) has been proposed in [18] to improve the trainability of the DCNNs. This model adopts residual connection to connect different layer, which can ignore some unimportant blocks in training automatically. This technique can solve some problems in traditional DCNNs. In order to make the DCNNs quickly locate the focal area of the image, inspired by the human visual mechanism, we propose a Visual Attentional Learning (VAL) block (shown in Figure 2(c)) based on the attentional mechanism. Specifically, we obtain the weight of channel and spatial position of the convolution feature by performing batch normalization on the feature map. This process can be expressed as:

$$\boldsymbol{w} = \emptyset(\boldsymbol{O}) \tag{1}$$

where $\boldsymbol{O} \in R^{H\times W\times C}$ means C-dimensional $H \times W$ feature map. In this paper, we adopt the features from last convolutional layer in each stage. The height of each feature is H and width is W; $\emptyset(\cdot)$ means batch normalization function. $\boldsymbol{w}$ is learned feature weight. Batch normalization function $\emptyset(\cdot)$ can be defined as:

$$\emptyset^S(\boldsymbol{O}) = \left\{m|m_{i,j}^c = \frac{e^{O_{i,j}^c}}{\sum_{i',j'} e^{O_{i,j}^c}}\right\} \tag{2}$$

$$\emptyset^C(\boldsymbol{O}) = \left\{m|m_{i,j}^c = \frac{e^{O_{i,j}^c}}{\sum_{c'} e^{O_{i,j}^{c'}}}\right\} \tag{3}$$

$$\emptyset^M(\boldsymbol{O}) = \left\{m|m_{i,j}^c = \sigma(C_{i,j}^c)\right\} \tag{4}$$

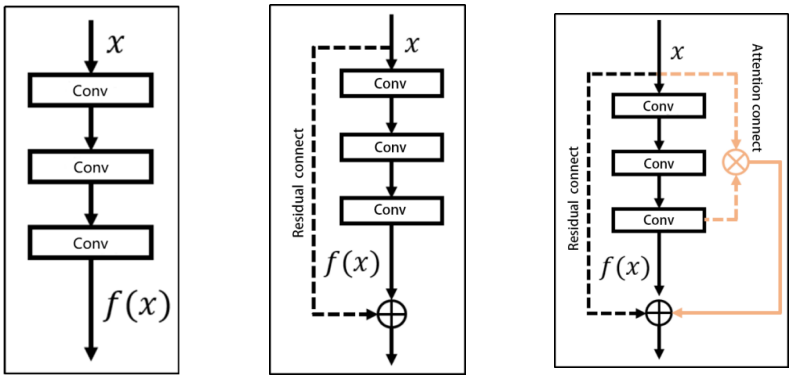
where (i,j) means the location in the feature map $\boldsymbol{O}$; c is the channel index of feature map; $\emptyset^S(\boldsymbol{O})$ is the attentional-driven block on spatial, which is used to learn feature weight on the spatial position; $\sigma(\cdot)$ is sigmoid function. $\emptyset^C(\boldsymbol{O})$ is the attentional-driven block on channel, which is used to learn feature weight of different channel dimension; $\emptyset^M(\boldsymbol{O})$ is the attention learning block that combines $\emptyset^S(\boldsymbol{O})$ and $\emptyset^C(\boldsymbol{O})$, which considers both spatial location information and channel information. In order to retain the advantages of the residual technique, we add the output of the attention strategy and the residual strategy as the final output of the block after weighting the convolution features. The process can be written as:

$$out^s = \emptyset^S(\boldsymbol{O}) * \boldsymbol{O} + \boldsymbol{O} + \mathrm{F}(\boldsymbol{O}) \tag{5}$$

$$out^c = \emptyset^C(\boldsymbol{O}) * \boldsymbol{O} + \mathrm{O} + \mathrm{F}(\mathrm{O}) \tag{6}$$

$$out^m = \emptyset^M(\boldsymbol{O}) * \boldsymbol{O} + \boldsymbol{O} + \mathrm{F}(\boldsymbol{O}) \tag{7}$$

where F(·) denotes residual connection. Based on the block defined above, the model not only can actively skip some unimportant features in the training process, but also can quickly locate some important channels and spatial positions by using attention mechanism. Therefore, the model can effectively alleviate the problems of insufficient training samples and small differences between samples of the same type in the flower Recognition task



(a) Tandem block (b) Residual block (c) Attentional-driven residual block

Figure 2. Block in different DCNNs models, (C) is our proposed block.

3.2. Attentional-driven residual network

We can structure any depth DCNNs models based on Attentional-driven residual block, but consider the local hardware and the scale of the dataset, we adopt ResNet14 and ResNet50 as the basic frameworks to construct Attentional-driven residual based version. The network structure is shown in Table 1. In this paper, we propose two novel methods, named VA-ResNet14 and VA-ResNet50, respectively. The input of two different depth model are both $224 \times 224 \times 3$ color jpg images, and then connected to the first deep learning block (convolution layer 1+), which consists by a 7×7 convolutional layer, a batch normalization layer, an activation layer and a maximum pooling layer. Then, we add the Attentional-driven residual block in the last layer of second (convolutional layer 2+), third (convolutional layer 3+), fourth (convolutional layer 4+) and fifth (convolutional layer 5+) stage. Not only that, but we retain the residual connection structure in the model (as shown in Figure 2 (c)). Finally, the model realizes the flower classification task though global average pooling and fully connected layer. The improved model is structurally identical to the original residual network. Since attentional-driven block has few parameters, the improved network will not increase the training burden. In addition, attentional-driven learning with residual connection can prove the performance of the model will not roll back. Even in the worst case, the residual connection can jump over the attentional learning block to make it down.

4. Experiments

4.1. Experiment setting

Based on the Attentional-driven residual network proposed above, we use Flowers 17 dataset to

evaluate its performance. By randomly dividing the augmented dataset according to proportion, we have 4760 images in training set with 280 images per class; 1360 images in validation set with 80 images per class and 680 images in testing set with 40 images per class. All of the flower images are two-dimensional color image in JPG format. The data for input need normalized by subtracting the mean value. The training process adopts Stochastic Gradient Descent (SGD) algorithm [15] to optimize the hinge loss function. The batch size is set to 128. The learning rate starts with 0.01, decreases to its 1/10 every 10,000 iterations, stops at 50,000 iterations. The weight decay parameter is 0.0005.

Table 1. Network structure of the attention learning model based ResNet.

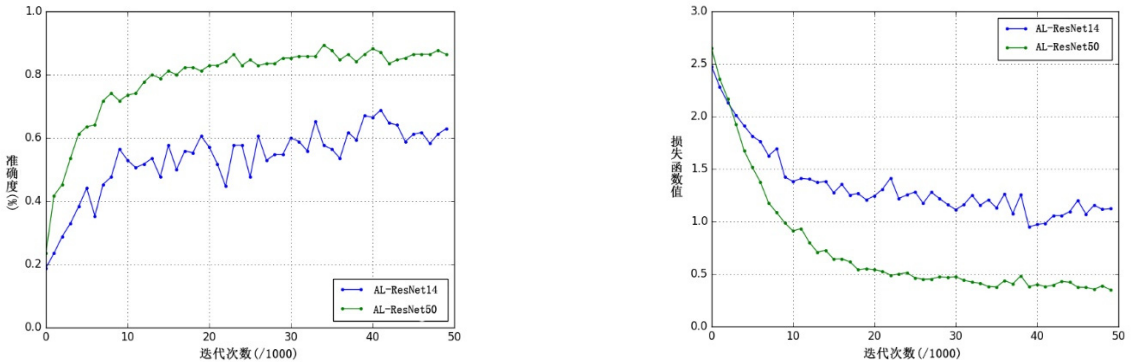
Layer	output size	ResNet14	VA-ResNet14	ResNet50	VA-ResNet50
Conv 1+	$112 \times 112 \times 64$	Conv, 7×7, stride 2	Conv, 7×7, stride 2	Conv, 7×7, stride 2	Conv, 7×7, stride 2
Conv 2+	$56 \times 56 \times 256$	Pool, 3×3, stride 2 Res block	Pool, 3×3, stride 2 Attentional Res block	Pool, 3×3, stride 2 [Res block] × 3	Pool, 3×3, stride 2 [Attentional Res block]×3
Conv 3+	$28 \times 28 \times 512$	$\begin{bmatrix}\text{Res block}\\ \text{stride 2}\end{bmatrix} \times 2$	$\begin{bmatrix}\text{Res block}\\ \text{stride 2}\\ \text{Attention connect}\end{bmatrix} \times 2$	$\begin{bmatrix}\text{Res block}\\ \text{stride 2}\end{bmatrix} \times 4$	$\begin{bmatrix}\text{Res block}\\ \text{stride 2}\\ \text{Attention connect}\end{bmatrix} \times 4$
Conv 4+	$14 \times 14 \times 1024$	$\begin{bmatrix}\text{Res block}\\ \text{stride 2}\end{bmatrix} \times 2$	$\begin{bmatrix}\text{Res block}\\ \text{stride 2}\\ \text{Attention connect}\end{bmatrix} \times 2$	$\begin{bmatrix}\text{Res block}\\ \text{stride 2}\end{bmatrix} \times 6$	$\begin{bmatrix}\text{Res block}\\ \text{stride 2}\\ \text{Attention connect}\end{bmatrix} \times 6$
Conv 5+	$7 \times 7 \times 2048$	$\begin{bmatrix}\text{Res block}\\ \text{stride 2}\end{bmatrix} \times 2$	$\begin{bmatrix}\text{Res block}\\ \text{stride 2}\\ \text{Attention connect}\end{bmatrix} \times 2$	$\begin{bmatrix}\text{Res block}\\ \text{stride 2}\end{bmatrix} \times 3$	$\begin{bmatrix}\text{Res block}\\ \text{stride 2}\\ \text{Attention connect}\end{bmatrix} \times 3$
Pool	$1 \times 1 \times 2048$	Global average pool			
Output	17	FC			

The experiment environment is Pytorch based on Python programming language. Pytorch as one of the most widely used framework in deep learning, has good scalability, modularity and high efficiency, which is very popular in the academic and industrial circles [19]. We implement all the algorithms in Think Station P320 workstation with 4 GTX 1080 Ti GPU to speed up image processing [20].

4.2. Results analysis

Based on experiment setting proposed above, we training the AL-ResNet14 and AL-ResNet50,

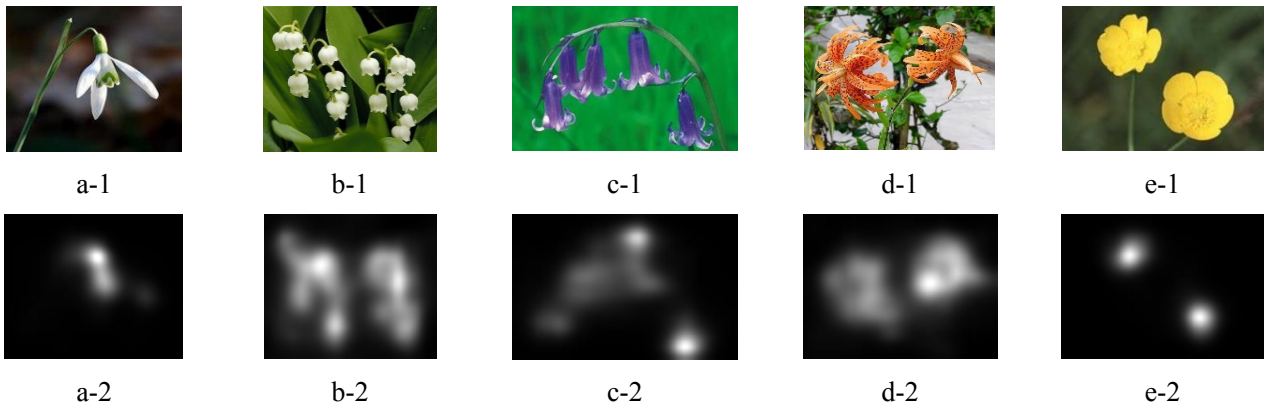
respectively. Figure 3 shows the curve of accuracy and objective loss in validation set, where the blue curve indicates the result in VA-ResNet14 and the green curve indicates the result in VA-ResNet50. We can find that the curve become placid after about 40,000 iterations, indicating that the algorithm has been converged. In addition, the accuracy on VA-ResNet50 is high that VA-ResNet14, but the objective loss is small, means VA-ResNet50 can fitting the data better. Besides, VA-ResNet14 is volatility higher that VA-ResNet50 in Figure 3, which is due to the number of VA-ResNet14 is less. In the case of same input data, the model with less parameters are hard to find the local optimal solution [21].



(a) accuracy on validation set (b) objective loss on validation set

Figure 3. Accuracy and loss of the validation in the training process.

To show the performance of Attentional-driven residual block, we provide the focus areas of flower images obtained by the attention learning in the first layer. As shown in the Figure 4, the brightly area is the model to focus on. We can see some interesting points as follow: 1) the focus area of attention is not continuous, but scattered into several bright spots.



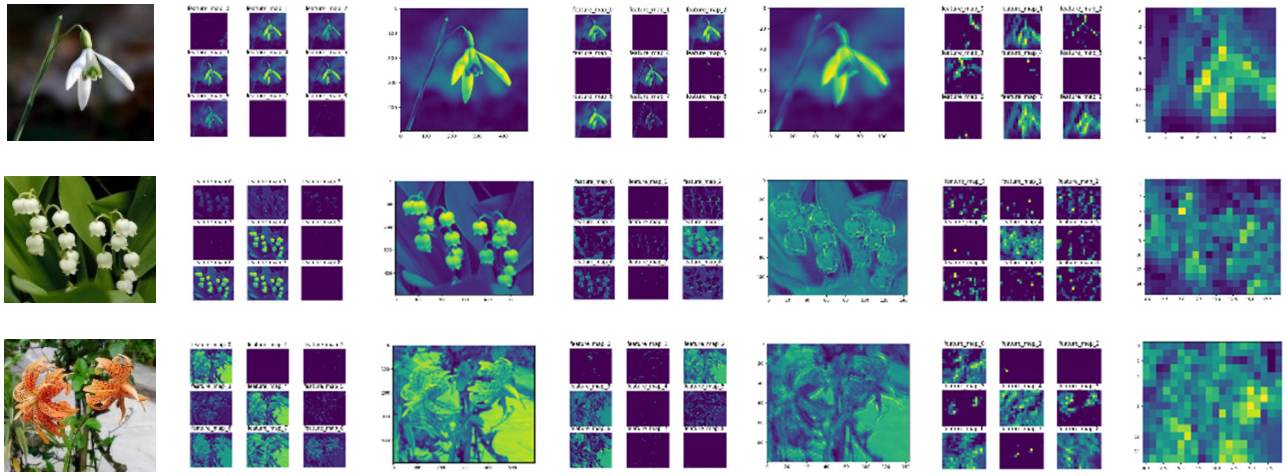
a-1 b-1 c-1 d-1 e-1

a-2 b-2 c-2 d-2 e-2

Figure 4. Focus areas of different varieties of flowers.

The brighter the area, the greater the role it plays in the classification, and the higher wright it corresponding to. It indicates that not all part of flowers plays an important role in flower recognition task. 2) Compare to original input image, the focus areas of flower are always corresponding to more colorful part in flower, indicating that the color information is the key point to discriminate the flower. Besides, since the attention mechanism puts more effort on flower, the noise in background

has no effect on flower recognition task, which bring robustness to model. Further, we visualize the convolution features from first to third layers of some flowers. Figure 5 shows the features visualization results of sunflower, snowdrop and tiger lily, respectively. From Figure 5 we can see the following conclusions. Firstly, the convolutional features learned by the shallow network are mainly understandable features such as texture and color, while the features learned by deep layers are more abstract, like outline or shape. 3) The feature from shallow layers are often high-resolution information, while the deep layers are more likely to extract some semantic information. Therefore, the resolution of images is gradually decreasing with the layer deep. In the classification process, the semantic information determines the image "what is", while the shallow features determine the discriminative information "where as" in the image.



(a) (b) First Layer (c) Second Layer (e) Third Layer

Figure 5. Feature visualization results of different flowers.

4.3. Method comparison

In order to further verify the effectiveness of the methods proposed in this paper, we compare our methods to some popular image classification techniques. We ensure all the parameters are consistent with the original text to guarantee the algorithms optimization. The results on the testing set are shown in Table 2.

Table 2. Accuracy comparison of different network models.

Method	VGGNet(16)	NIN	GoogLeNet	Inception V3	ResNet14	ResNet50	VA-ResNet14	VA-ResNet50
Accuacy	63.1%	64.2%	65.8%	66.9%	67.7%	81.3%	69.4%	85.7%

Comparing our method with VGGNet [22], Network In Network [23], GoogLeNet [24], and Inception V3 [25], we can find that the method proposed in this paper has higher accuracy. Specifically, VA-ResNet14 and VA-ResNet50 have accuracy improvement of 1.7 and 3.6% than ResNet14 [18] and ResNet50 [18], respectively. This indicates that the proposed method has good universality. VA can still improve the model performance, even on the DCNN model with a strong

presentation ability. Also, it can be found that ResNet with VA blocks shares very higher accuracy as compared to VGGNet [22], Network In Network [23], GoogLeNet [24], and Inception V3 [25].

5. Conclusions

In this paper, we propose a novel Attentional-driven residual network model for flower recognition. By adding an attention connection to each residual block, the model can learn from different channel features and different spatial dimensions, and at the same time, can maintain the capability of few-shot learning to compensate training samples insufficient. In order to verify the feasibility and effectiveness of the methods proposed in this paper, we take the experiments on Flowers 17 dataset. The experiments show that our method can achieve the accuracy of 85.7%, which is higher than the existing image classification methods without introducing additional training parameters. Although the methods proposed in this paper is initially designed for flower recognition, it has strong scalability and practicability that can be easily applied to other object recognition tasks, such as terrain recognition, farmland recognition, and forest recognition on remote sensing images.

In addition, our future work will focus on the following aspects. 1) Expand the flower database. Not only we expand the number of flower varieties, but also expand the number of images in each species. 2) Since our methods are supervised learning model, which need using a lot of labeled data in training process, one of our future projects is to combine with some advanced technique, like semi-supervised learning, one/few-shot learning. 3) Another project will focus on the transfer learning and data generation technique based on natural image datasets to improve the generalization ability and robustness of the model.

Acknowledgments

This work was supported by Startup Foundation for Introducing Talent of Nanjing University of Information Science and Technology (Grant No.2019r030).

Conflict of interest

The authors declared that they have no conflicts of interest to this work. We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted

References

1. D. R. Pereira, J. P. Papa, G. F. R. Saraiva, G. M. Souza, Automatic classification of plant electrophysiological responses to environmental stimuli using machine learning and interval arithmetic, *Comput. Electron. Agric.*, **145** (2018), 35–42.
2. Q. L. Ye, Z. Li, L. Fu, Z. Zhang, W. Yang, G. Yang, Nonpeaked Discriminant Analysis Data Representation, *IEEE Trans. Neural Networks Learn. Syst.*, **30** (2019), 3818–3832.

3. Q. L. Ye, J. Yang, F. Liu, C. Zhao, N. Ye, T. Yin, L1-Norm Distance Linear Discriminant Analysis Based on an Effective Iterative Algorithm, *IEEE Trans. Circuits Syst. Video Technol.*, **28** (2018), 114–129.
4. L. Fu, Z. Li, Q. L. Ye, H. Yin, Q. Liu, X. Chen, et al., Learning Robust Discriminant Subspace Based on Joint L2,p- and L2,s-Norm Distance Metrics, *IEEE Trans. Neural Networks Learn. Syst.*, 2020, forthcoming.
5. L. Fu, D. Zhang, Q. L. Ye, Recurrent Thrifty Attention Network for Remote Sensing Scene Recognition, *IEEE Trans. Geosci. Remote Sens.*, 2020, forthcoming.
6. C. Wachinger, M. Reuter, T. Klein, DeepNAT: Deep convolutional neural network for segmenting neuroanatomy, *NeuroImage*, **170** (2018), 434–445.
7. Y. Cheng, L. Fu, P. Luo, Q. Ye, F. Liu, W. Zhu, Multi-view generalized support vector machine via mining the inherent relationship between views with applications to face and fire smoke recognition, *Knowl. Based Syst.*, **210** (2020), 106488.
8. Y. Chen, H. Yin, Q. L. Ye, P. Huang, L. Fu, Z. Yang, Improved multi-view GEPSVM via Inter-View Difference Maximization and Intra-view Agreement Minimization, *Neural Networks*, **125** (2020), 313–329.
9. Q. L. Ye, H. Zhao, Z. Li, X. Yang, S. Gao, T. Yin, et al., L1-norm Distance Minimization Based Fast Robust Twin Support Vector k-plane clustering, *IEEE Trans. Neural Networks Learn. Syst.*, **29** (2018), 4494–4503.
10. H. Zhu, Q. Liu, Y. Qi, X. Huang, F. Jiang, S. Zhang, Plant identification based on very deep convolutional neural networks, *Multimedia Tools Appl.*, **77** (2018), 29779–29797.
11. Q. Ye, D. Xu, D. Zhang, Remote sensing image classification based on deep learning features and support vector machine, *J. For. Eng.*, **4** (2019), 20961359.
12. Y. Liu, X. Zhou, Z. Hu, Y. Yu, Y. Yang, C. Xu, Wood defect recognition based on optimized convolution neural network algorithm, *J. For. Eng.*, **4** (2019), 115–120.
13. M. Cıbuk, U. Budak, Y. Guo, M. C. Ince, A. Sengur, Efficient deep features selections and classification for flower species recognition, *Measurement*, **137** (2019), 7–13.
14. S. Zhang, J. Yang, B. Schiele, *Occluded Pedestrian Detection Through Guided Attention in CNNs*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.
15. M. E. Nilsback, A. Zisserman, *A Visual Vocabulary for Flower Classification*, 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2006.
16. J. Zhang, K. Shao, X. Luo, Small sample image recognition using improved Convolutional Neural Network, *J. Visual Commun. Image Representation,* **55** (2018), 640–647.
17. K. Li, M. Zhang, Z. Yang, B. Lyu, Classification for decorative papers of wood-based panels using color and glossiness parameters in combination with neural network method, *J. For. Eng.*, **3** (2018), 16–20.
18. K. He, X. Zhang, S. Ren, J. Sun, *Deep Residual Learning for Image Recognition,* Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.
19. M. Ravanelli, T. Parcollet, Y. Bengio, *The PyTorch-Kaldi Speech Recognition Toolkit*, ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2019.
20. B. Kim, C. Oh, Y. Yi, D. Kim, GPU-Accelerated Boussinesq Model Using Compute Unified Device Architecture FORTRAN, *J. Coastal Res.*, **85** (2018), 1176–1180.

21. G. Cheng, Z. Li, J. Han, X. Yao, L. Gao, Exploring hierarchical convolutional features for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.*, **56** (2018), 6712–6722.
22. T. Sercu, C. Puhrsch, B. Kingsbury, Y. LeCun, *Very Deep Multilingual Convolutional Neural Networks for LVCSR*, 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2016.
23. M. Lin, Q. Chen, S. Yan, Network in network, preprint, arXiv:1312.4400,
24. K. He, X. Zhang, S. Ren, J. Sun, *Delving Deep into Rectifiers: Surpassing Human-level Performance on Imagenet Classification*, Proceedings of the IEEE international conference on computer vision, 2015.
25. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, *Rethinking the Inception Architecture for Computer Vision*, Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.

AIMS Press