Advances in Artificial Intelligence, Big Data and Algorithms
G. Grigoras and P. Lorenz (Eds.)

doi:10.3233/FAIA230862

Fine-Grained Classification of Wild Fungi Based on Attention Residual Mechanism

Shuoyi WEN[1], Li YANG, Hailong DUAN and Tingting ZHANG
Tianjin Key Laboratory of Information Sensing and Intelligent Control, Tianjin University of Technology and Education, Tianjing, China

Abstract. Identifying wild mushroom species are an important way to prevent poisoning from consuming toxic wild mushrooms. Therefore, an attention-based method for fine-grained classification of wild mushrooms is proposed, in which an attention mechanism is incorporated and a new network model structure is constructed in combination with a residual module. Firstly, nine diseased wild mushroom image samples were collected, and in order to make the model have better generalization ability, the images were pre-processed so that the number of samples reached 7200, and the experimental results showed that the improved deep residual network model achieved about 99.12% accuracy in classifying and recognizing the established wild mushroom database, and the improved neural network had a great improvement in classification accuracy.

Keywords. attention; fine-grained picture; classification; wild mushrooms

1. Introduction

In Yunnan Province, there were 273 wild mushroom poisoning incidents between May and July 20, 2020, in which 12 people died. If there are so many risks in consuming wild mushrooms why do people still enjoy it? First of all, in terms of quantity, the "Guang mu Shu" written by Pan Zhiheng in the Ming Dynasty recorded 119 kinds of actual fungi, and there are currently 250 kinds of wild fungi in Yunnan, accounting for about one-half of the total number of edible mushrooms in the world; secondly, it is purely natural, Yunnan wild mushrooms can grow freely because of their unique geographical location, environment and climate; thirdly, it is delicious, Yunnan wild mushrooms have the characteristics of fresh, thick, fat and tender; finally, it is Lastly, the medicinal value of the mushrooms is extensive, with some health benefits and some mushrooms can be used as medicine1. Due to the uneven distribution and variety of wild mushrooms, it is difficult for people to distinguish them from edible ones, so poisoning often occurs and an efficient identification method is urgently needed.

The biggest difference between generic image classification and fine-grained image classification is that generic image classification distinguishes, for example, the classification of dogs and cats and so on broad categories, for fine-grained image classification is the recognition of different breeds in a category, for example, whether the dog category is the classification of Tibetan mastiff, Golden or Chenery, the main difficulty of fine-grained classification is that the gap between classes of different classes

[1]Corresponding Author: Shuoyi WEN, Tianjin Key Laboratory of Information Sensing and Intelligent Control, Tianjin University of Technology and Education; e-mail: 1315282586@qq.com

is small, and the gap between classes of the same class is large. To address the problem of small information disparity in fine-grained image classification, Li et al. proposed a fine-grained image classification model of bilinear aggregated residual attention[1] which has a correct classification rate of 87.9%, 92.9%, and 94.7% for three fine-grained images of CUB-200-2011, FGVC-Aircraft, and Stanford Cars, respectively[2]. For fine-grained classification of indistinguishable wood, Dai et al. proposed a new automatic classification algorithm based on deep convolutional neural network for fine-grained image recognition, which is enough to achieve more accurate data classification, and the results show that the model has 88.36% accuracy, which has high practical value and can effectively improve the classification accuracy of wood species and provide a basis for fast classification of wood species[3].For the classification of wild mushrooms, Zhang et al. first designed a multi-scale feature guidance (MSFG) module, which improved the accuracy of the proposed improved model by 8.10 percentage points respectively compared to the previous maximum4. Separate datasets of wild mushrooms were built, and the accuracy also reached 96.32% compared to the above methods in terms of data volume and accuracy[4].

2. Material and Method

2.1. Introduction to the Data Set

In this paper, nine species of wild mushroom images were used as the main object of study and the corresponding species of wild mushrooms were collected to ensure the relative richness and perfection of the data. A total of 9 species based on online selected and field photographed images, taking a training set: test set = 9:1, with a total of 7200 images. Containing larch mushroom, umbrella poisonous mushroom, boletus, filamentous mushroom, pink fold mushroom, scarlet wet umbrella, milk mushroom, red mushroom, and milk boletus. As shown in Figure 1.

Figure 1. Wild mushroom dataset

2.2. Data Pre-Processing

Pre-processing of the images, in order to improve the image quality, the images are firstly cleaned to remove the damaged and problematic images, based on which the original images are inverted, the brightness is converted to 0.7~1.3 and Gaussian noise is added with a variance of 0.02. By data enhancement of the images, the sampling quality of the images is improved and networks with stronger generalization capability can be obtained, which are better adapted to apply to different scenes.

2.3. Methodology

1) Convolutional neural networks

Convolutional Neural Networks[5] The basic process can be divided into a combination of convolution layer, nonlinear mapping (ReLU layer), pooling layer, fully connected layer, and output. As in Figure 2 each layer will have a different role. The convolution layer has weight sharing in it and automatically extracts features; the nonlinear mapping allows for a better fit of the model. The pooling layer filters the information and reduces the dimensionality of the features to reduce the computation and prevent overfitting; the fully connected layer is a long vector that will be at the end of the convolutional network and combined with the entire connected layer output layer, and finally the classification result is obtained.

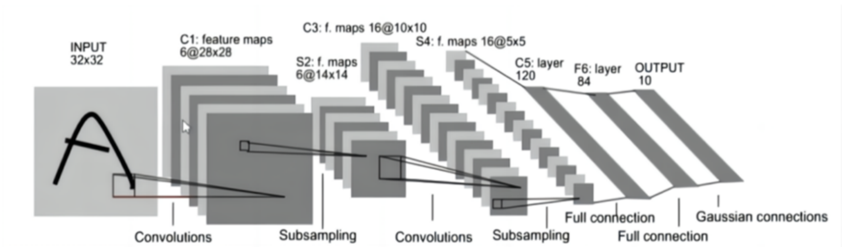

Figure 2. Convolutional neural network

The convolution layer allows the local features of the image to be extracted and then a new layer is obtained using the excitation function. The formula for this convolution layer is.(l is the serial number layer. M_j is the set of input features. x_j^l is the first I layer of the j neuron output. x_j^{l-1} is the first I layer I neuron input)

$$x_j^l = f \sum_{i \in M_j} x_i^{l-1} k_{ij}^l + b_j^l \tag{1}$$

The nonlinear activation function uses a modified linear unit (ReLU function) x is the function independent variable, and $f(x)$ is the ReLU function)

$$f(x) = \begin{cases} 0(x<0) \\ x(x>0) \end{cases} \tag{2}$$

The pooling layer reduces the dimensionality of the feature map and reduces the likelihood of overfitting. Pooling is calculated as.($f_{down}\left(x_i^{l-1}\right)$ is the downsampling function)

$$x_j^l = f_{down}\left(x_i^{l-1}\right) \tag{3}$$

The fully connected layer is located after the pooling layer and further dimensionality reduction is applied to the feature and this feature is fed into the SoftMax classer. To reduce the loss function, a gradient descent method is used. The loss function equation.(w is the weight. b is the bias term. N is the total number of training samples. C is the number of training sample categories. I is the indicator function. j is the training sample category. $\hat{y}_i$ is the sample training expectation. p_i^j is the predicted probability)

$$L(w,b) = -\sum_{i=1}^{N}\sum_{j=1}^{C} I\{\hat{y}_i = j\} \lg p_i^j \tag{4}$$

2) Residuals module

In the development and application of neural networks, for the network, the more hidden layers, the deeper the model, he should work better, but some models get worse the more they are trained. In 2016 when residual networks were proposed, theoretically, the more layers of stacked neural networks should improve the model accuracy, but at first, as the number of layers of the model increases, the accuracy of the model will reach saturation, and if in increasing the number of network layers then the network starts to degrade, the reason being that the deeper networks are too difficult to train resulting in. The concept of residual is the deviation of the prediction from the actual data on probability statistics and the residual network is fitted to it. While residuals are used to design to solve the degradation problem, its also solves the gradient problem and moreover makes the performance of the network improved as well. The problem of gradient disappearance or gradient explosion due to deepening of layers is cleverly solved [6][7].

The introduction of residual models is a key aspect in the development of convolutional neural networks. A residual network is composed of a series of residual blocks[8] as in Figure 3. A residual block can be represented as $x_{L+1} = x_L + f(x_L, w_l)$. The residual block consists of a direct mapping part and a residual part. $f(x_L, w_l)$ It is the residual part that generally consists of two or three convolution operations. In the residual block, the input can be propagated forward more quickly over multiple levels of the data line. The model input is divided into two paths the rightmost line is a jump connection directly passing the input as is to the output, inside is a two-layer neural network with an activation function added between.

The residual network was developed based on the VGG network and the key to its success is the short-circuiting mechanism by adding the residual learning module. Which has many residual blocks stacked, thus solving the network degradation problem, with the deepening of the model complexity, from ResNet18, ResNet34, ResNet50, ResNet101 developed to ResNe152 and so on, Figure 4 ResNet50 detailed structure[9], by the backbone and there are four major groups of blocks are composed of 3, 4, 6, 3 small block, each small black contains three convolutions inside, in addition to the very beginning of this network has a separate 7×7 convolutional layer, plus the last fully connected layer for a total of 50 layers. This algorithm reduces the computational effort substantially while ensuring the original backbone information, and also can effectively reduce the computational effort of the model.

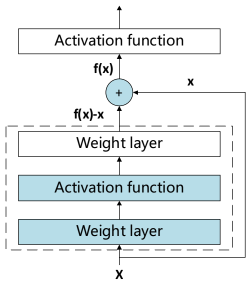

Figure 3. Residual block

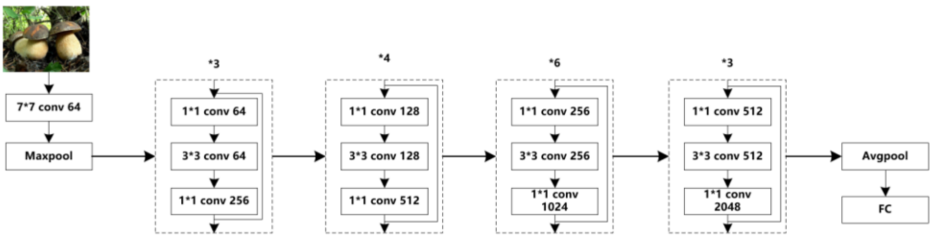

Figure 4. Res net-50 structure diagram

3) Attention module

Humans have a wide field of vision, but the focus is only to a small area, or a point, this is called attention, attention is a survival mechanism for humans in the long-term development, in the current computer computing power resource limitations, attention mechanism is absolutely a necessary means to improve efficiency, focus attention on useful information entropy, do not spend time in the noise.SE (Squeeze-and-Excitation) module also known as channel attention[10], as Figure 5 input over a feature map after a convolution to get a new feature map, this feature map contains c feature channels, each channel above the width and height of HW, respectively, after a branch above the normal flow, the branch contains three main parts, the first compression (Squeeze) can take the current feature map for each channel, are compressed. For example, the current channel is a two-dimensional vector of $h \times w$, and this two-dimensional vector is globally pooled into a real value, which is a global perceptual field, and then the final output is a new feature map of $1 \times 1 \times C$ [11][12], the number of channels of this new feature map and the number of channels of the original feature map is the same ex is an incentive branch, in learning the importance of each channel, scale This is rescaled, and the learned importance of each channel is multiplied by the original feature map and the main channel to get a new one.

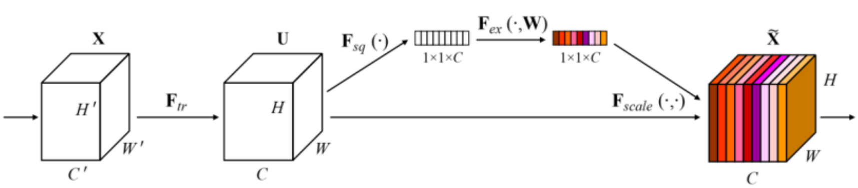

Figure 5. Channel attention Figure

Spatial attention module: first the average pooling and maximum pooling along the channel axis is calculated, then the resulting feature information is input into the convolution, and finally the spatial attention feature is obtained by an activation function. On the basis of channel attention, spatial attention is the most important location of the key information, and thus the feature map.

4） Model construction

This study uses ResNet50 as the network model, which enhances the depth of the model while avoiding gradient disappearance. The overall structure diagram is shown in Figure 6.

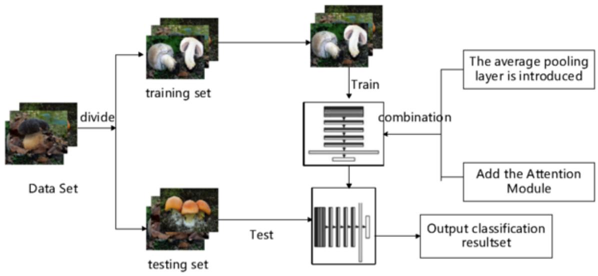

Figure 6. Overall structure diagram

For the improved residual network, the convolutional layer of the XY path loses 75% of the input feature information by using a change in down sampling. Since the step size of the previous convolution is 2 and the convolution kernel is 1×1, both paths lose a lot of input feature information. As shown in Figure 7: add a 2×2 average cellularization before the convolution of path X to reduce the lost feature values. As shown in Figure7 first method: add a 2×2 average pooling layer in front of the convolution of path X to reduce the lost feature values, second method: add a 2×2 average pooling layer on both paths XY.

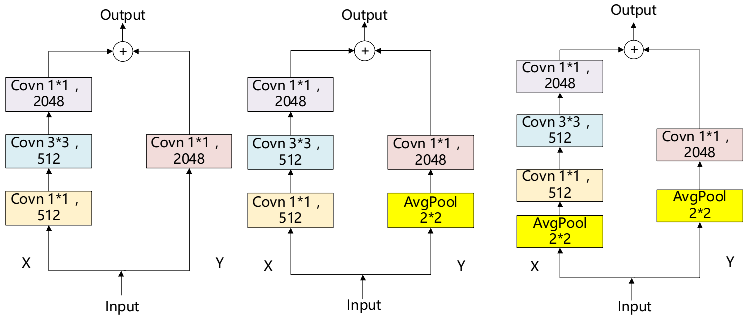

Figure 7. Adding down sampling

The convolutional ResNet50 is added as extracted features with iterative attention, and the features extracted by ResNet50 are further extracted features inside the iterative attention module, which combined with the original features can be more accurate to improve the ability of filtering images and thus improve the accuracy. x is passed through spatial attention and channel attention to get two different weights, x and channel attention are added together, x and spatial attention are multiplied, and the result is obtained after adding with the above. The feature extraction of wild fungus feature maps is achieved by adding attention module in each level of the modified ResNet50 as in Figure 8, adding attention in both spatial and channel dimensions and then multiplying it with the input feature map.

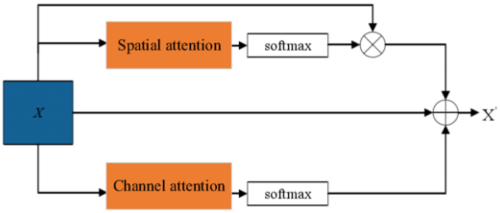

Figure 8. Adding the attention structure

3. Results and Analysis

First in the residual network series for training, from which to find the most suitable for improving a, through the table 1, can be derived as the number of layers of the residual network continues to increase, the depth continues to deepen, the accuracy of network recognition, gradually improve. Resnet50 resnet101 Resnet152 achieves approximately the same accuracy and ResNet101, Res net152 network is deeper the number of parameters is large, so ResNet50 as the object of study for subsequent improvement operations.

Table 1. Comparison of accuracy rates

models	ResNet18	ResNet34	ResNet50	ResNet101	ResNet152
Accuracy rate/%	70.02	75.45	82.53	83.22	83.15

Based on the above mentioned methods for residual network improvement, 200 rounds of tests were conducted to obtain the results: method 1 specified paths to join the average pooling layer, method 2 two paths to join the average pooling layer, the results are shown in Figure 9. The results of method two are better.

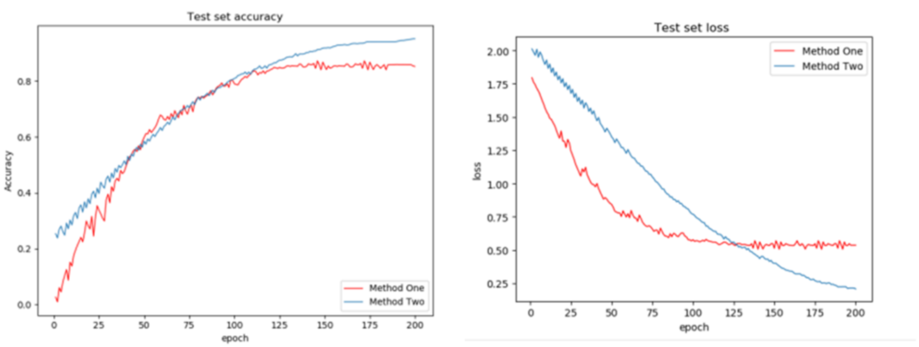

Figure 9. Method 1 Method 2 accuracy and loss rates

The results of the attention mechanism network experiment are shown in Figure 10, where the network is optimized while continuing to increase channel attention and spatial attention. The accuracy and loss rate curves are obtained to start converging at about 60 times, better than the original model at about 99.12% accuracy and loss at about 0.03, better improving the recognition and reducing the loss rate.

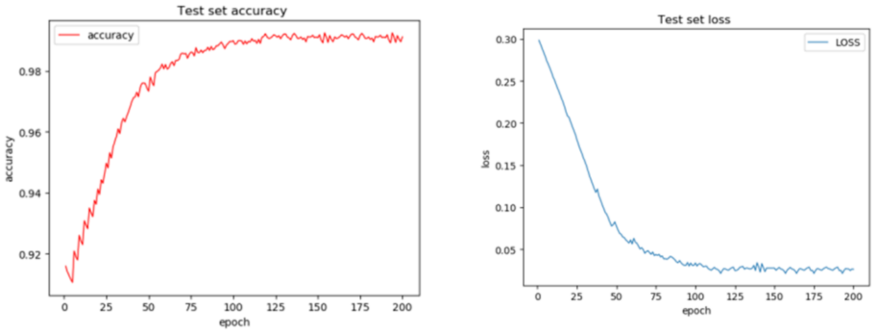

Figure 10. Accuracy of adding attention mechanism

4. Conclusion

For wild mushroom fine-grained image classification, a method based on attention wild mushroom fine-grained classification, which incorporates an attention mechanism and combines the residual module to construct a new network model structure Experiments show that the method has good recognition ability with about 99% correctness of recognition. However, there are still many problems that need further improvement, such as expanding the variety of wild mushroom datasets, etc.

References

[1] Yang Zhuliang. A brief discussion on wild mushroom resources and their utilization in Yunnan [J]. Journal of Natural Resources, 2002(04):463-469.
[2] Christian S, Sergey I, Vincent V, et al. Inception-v4, inception-resnet and the impact of residua l connections on learning[J]. ar Xiv preprint ar Xiv:1602.07261, 2016.
[3] DAI Tianhong,XIE Qiancheng,HUANG Jianping,SUN Chunxue,CONG Shijie,HUANG Xinwang, LI Kexin. A weakly supervised fine-grained deep network approach for wood classification[J]. J ournal of Southwestern University (Natural Science Edition),2022,44(10):161-172.DOI:10.13718/j. cnki.xdzk.2022.10.017.
[4] Zhang ZG,Yu PF,Li HY,Li HS. Fine-grained wild mushroom image recognition based on multi-scale feature guidance[J]. Advances in Laser and Optoelectronics,2022,59(12):192-201.
[5] Z. Li, F. Liu, W. Yang, S. Peng and J. Zhou, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," in IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 12, pp. 6999-7019, Dec. 2022, doi: 10.1109/TNNLS.2021.3084827.
[6] Li Lipeng,Shi Fipeng,Tian Wenbo,Chen Lei. A wild plant image recognition method based on r esidual network and migration learning[J]. Radio Engineering,2021,51(09):857-863.Gupta Arpita,P awade Pratik,Balakrishnan Ramadoss. Deep Residual Network and Transfer Learning-based Perso n Re- Identification[J]. Intelligent Systems with Applications,2022,16.
[7] Tang Buchen,Liang Ruiyu,Xie Yue,Huang Yulei,Wang Qingyun. Acoustic scene classification m ethod based on attention mechanism and two-path deep residual network[P]. Jiangsu Province: CN111754988B,2022-08-16.
[8] Huang Xuan,Sun Han,Lin Bosheng,Yin Mingjun,Yang Zhijun. A fiber classification method base d on improved ResNet-50 residual network[J]. Journal of Xi'an University of Engineering,2022,3 6(04):19-25.DOI:10.13338/j.issn.1674-649x.2022.04.003.

[9] Zhang D., Cai X. S., Jiang S. D., Zuo H. X., Xu Z., Zheng Zheng Y.. Waste classification sy stem based on Resnet50[J]. Computer Programming Skills and Maintenance,2022(09):24-27.DOI: 10.16184/j.cnki.comprg.2022.09.052.

[10] Sochor J, Jakub p, Herout A. Boxcars: 3D boxes as CNN input for improved fine-grained vehi cle recognition[C]//Conference on Computer Vision and Pattern Recognition. 2016(5): 3006-3015.

[11] Qian Jia-Xin,Yu Peng-Fei,Li Hai-Yan,Li Hong-Song. Fine-grained classification of wild mushroo ms based on feature fusion and attention[J/OL]. Advances in Lasers and Optoelectronics:1-16[20 22-11-01]. http://kns.cnki.net/kcms/detail/31.1690.TN.20220713.1913.571.html

[12] Li S H, Liu Y H, Zhang R F. Fine-Grained Image Classification Based on Multi-Scale Feature Fusion[J]. Laser & Optoelectronics Progress, 2020, 57(12): 121002.