Rice Disease and Insect Pest Identification Algorithm with Shuffle Attention Mechanism

Yuliang Gao ^{a,b,*}, Lifeng Zhang, Bin Li

^a Kyushu Institute of Technology, Kitakyushu, Fukuoka 804-0011, Japan

^b Yangzhou University, Yangzhou, Jiangsu 225000, China

* Corresponding Author: gao.yuliang101@mail.kyutech.jp

Abstract

This paper presents an identification algorithm of rice diseases and insect pests based on convolutional neural network. Aiming at the problems of large amount of parameters and low accuracy of traditional algorithm. The traditional convolution neural network Inception-ResNetV2 and the shuffle attention module are combined. It achieves the effects of greatly reducing parameters and obviously improving accuracy. Compared with traditional models ResNet50 and Inception-ResNetV2, the highest accuracy reaches 95.24%, the parameters in the self-made data set decreased by 47.6% and 34.02%, and the accuracy increased by 1.49% and 0.9%, respectively.

keywords: Incidence rice disease and insect pest identification, Inception-ResNetV2, shuffle attention

1. Introduction

With the development of smart agriculture, the combination of information technology and traditional agriculture has achieved good results. For the recognition of common rice diseases and insect pests, the image recognition method is mainly used at present, and the deep learning network is always a hot research direction. Deep learning network has the advantages of good effect, strong generalization ability and easy migration. In recent years, excellent backbone deep learning networks continue to appear, such as ResNet⁽¹⁾ of He, GoogleNet⁽²⁾ of Ballester, Inception⁽³⁾ of Szegedy, VGG⁽⁵⁾ of Sengupta, etc. which are widely used in image classification and recognition. At the level of training methods, the emergence of transfer learning, pre training model and other methods also make the training of high-quality model more simple. In addition, the emergence of attention mechanism to improve the performance of backbone network

greatly improves the accuracy of classification tasks, such as Woo's CBAM,⁽⁸⁾ Hu's SENet,⁽⁹⁾ Wu's SKNet,⁽¹⁰⁾ etc. With these methods, the top 1 accuracy of ImageNet classification task is improved from 71.93% of resnet50 to 79.81% of SKNet.

For the identification of rice diseases and insect pests, Lu⁽¹¹⁾ used the traditional deep convolution neural network to learn and identify 10 kinds of rice diseases and pests, and the accuracy reached 95.48%. Liang⁽¹²⁾ also used the deep convolution network to identify rice blast, and the accuracy reached 95%.

The epidemic raged, many countries stop exporting chemical fertilizers and grain, and the risk of global food crisis is increasing. The security requirements for rice planting are higher and higher. Traditional deep learning network structure has low accuracy and large number of parameters. It is more and more difficult to meet the practical requirements. So it is very important to design a recognition algorithm with small parameters and high accuracy. The above research also provides a reference and feasibility for the algorithm design of this paper.

ResNet50 and Inception are classical traditional algorithms in the field of image classification. Residual network is first proposed in ResNet model to solve the problem of gradient explosion or disappearance. The Inception model can improve the network width and performance by adding different convolution kernels. These two algorithms are classic traditional algorithms for classification and recognition tasks. In order to solve the problems of low accuracy and large amount of parameters of these two traditional algorithms, this paper combines the Inception-ResNetV2⁽³⁾ model with shuffle attention⁽¹³⁾, which can significantly improve the accuracy. For the task of rice pest identification, SA-Light-Inception-ResNetV2(SLIR) algorithm was designed to meet the requirements. Compared with the traditional algorithm,

in the case of using fewer parameters, higher accuracy is obtained.

2. SLIR Rice Pest Identification Network

2.1 Inception-ResNetv2 module

Inception is a convolutional neural network model proposed by Google in 2015. Its way is to widen the network of the neural network, divide the neural network into several branches, and convolute the input features to extract features, so as to get multi-scale information. Finally, the features are superimposed to get the output features. The idea of residual network ResNet is to skip the middle parameter layer and directly output it to the output layer. That is, if the input is x and the Inception layer is $f(x)$, then the output is $y = f(x) + x$. The idea of Inception-ResNetV2 is to combine the traditional Inception module and ResNet module. Inception-ResNetV2 module has six structures, as shown from Fig.1 to Fig.6. Stem extract image features. Inception-ResNetV2-A, B and C extract features of different dimensions by superposition of different convolution kernels. Reduction-A and B change the feature dimension through convolution network to input to the next layer.

2.2 Shuffle Attention

Shuffle attention mechanism is based on the traditional spatial attention mechanism and channel attention mechanism, adding the idea of group processing and re-combination. The input features are divided into groups and blocks, and then the spatial attention and channel attention weights are assigned. The flow chart of the algorithm is shown in Fig.7: grouping attention first groups the input in the channel dimension according to the parameter groups, and then divides each group into two parts X_{k1} and X_{k2} in the channel dimension and then attach the weight of channel attention and spatial attention respectively.

Channel attention, as shown in equation.1. X_{k1} first passes through the avgpool function to get the channel statistics data, then enhances the feature representation through the fuse linear function, and finally multiplies the original value through the sigmoid function to embed the global information. Through the channel attention mechanism composed of these three functions, The X_{k1}' with channel attention weight is obtained. It can better reflects the "what" attribute of the feature, that is, what needs to be recognized;

Fig. 1. Stem

$$X_{k1}' = \sigma(F_C(F_A(X_{k1}))) \cdot X_{k1} = \sigma(W_1 S_1 + b_1) \cdot X_{k1} \quad (1)$$

Where: F_C are linear functions, σ is the sigmoid activation function, F_A is avgpool function, S_1 is the feature after average pooling. S_1 and W_1 are linear function parameters.

Spatial attention, as shown in equation.2. X_{k2} is first normalized by groupnorm function to obtain the spatial statistical data, then enhanced by fuse linear function, and finally multiplied by the original value after activation by sigmoid function to embed the global information. Through the spatial

Fig. 2. Inception-ResNetV2-A

Fig. 4. Inception-ResNetV2-C

Fig. 3. Inception-ResNetV2-B

Fig. 5. Reduction-A

Fig. 6. Reduction-B

attention mechanism composed of these three functions, the feature X_{k2}' with spatial attention weight is obtained, let the feature better reflect the "where" attribute of the feature, that is, where the object needs to be recognized.

$$X_{k2}' = \sigma(F_{C2}(F_{GN}(X_{k2}))) \cdot X_{k2} = \sigma(W_2 S_2 + b_2) \cdot X_{k2} \quad (2)$$

Where: F_{C2} are linear functions, σ is the sigmoid activation function, F_{GN} is GroupNorm function, S_2 is the feature after GroupNorm. S_2 and W_2 are linear function parameters.

2.3 SA-Light-Inception-ResNetV2(SLIR)

As shown in Fig.8, the algorithm takes into account the advantages of the Inception-ResNetV2 module in multi-

dimensional extraction of image features, and the advantages of grouping attention mechanism in the attention weight assignment of overlay features. The Inception-ResNetV2 module is combined with the group attention mechanism, and then the attention weight is attached to the feature by using the group attention mechanism after each structure of Inception-ResNetV2. The SLIR model designed in this paper extracts features by three Inception-ResNetV2 modules, assigns attention weight by grouping attention mechanism, and changes feature dimensions by reduction to output the next layer. The input of the whole model is dimension 299 × 299 × 3 samples of pest images are first divided into two convolutional blocks composed of three convolution cores, and the two convolution blocks are connected by pool layer. The output of the previous layer is the input of the next layer, and so on. Then, it passes through one packet attention module, and then passes through three inception ResNetV2 modules and full connection layer. Each Inceptions-ResNetV2 module adds a group attention mechanism. Finally, the output is completed by a full connection layer and activation function.

3. Experiment Environment Setting And Result Display

In this paper, Pytorch is used as a deep learning framework, and the hardware is set with GTX1080, AMD Ryzen 53500 CPU and 16GB memory. Using SGD as the optimization algorithm, the learning rate is 0.01, and the

Fig. 7. Shuffle Attention

Fig. 8. SLIR

Fig. 9. Examples of Incidence rice disease and insect pest images

Fig. 10. Training accuracy

decline is 0.0005. The learning rate is adjusted once every 50 rounds. The training round is 150 rounds. Under this environment parameter, the accuracy rate (total number of correct recognition of test set/total number of pictures of test set) is taken as the evaluation index, and the algorithm of Inception-ResNetV2 and ResNet50 is taken as the Baseline. The SLIR(SLIR-without-SA) model without adding group attention mechanism is run to test the effect of group attention mechanism.

The accuracy changes of SLIR model and traditional model in training are shown in Fig.10. Among them, accuracy represents the recognition accuracy on the test set, and epoch represents the training round.

The parameters of different models and the highest accuracy after training are shown in Tab.1.

The attention weights of different models to pest pictures are shown in Fig.9. Compared with the original picture, the redder the area, the greater the attention weight of the model and the greater the impact on the final classification results.

3.1 Analysis of Results

It can be seen from Fig.10 that all the models are close to convergence after 50 rounds. In the training process, the ac-

Table 1. Model parameters number and Accuracy

Model name	Parameters Number	Accuracy
SLIR	19975519	95.24%
SLIR-without-SA	19975333	94.57%
Inception-ResNetV2	30277413	94.34%
ResNet50	38089797	93.75%

Fig. 11. Attention Heat Maps

curacy of the SLIR model proposed in this paper is obviously higher than that of the traditional model. And the highest accuracy is shown in Tab.1. Compared with the traditional models ResNet50 and Inception-ResNetV2, the SLIR model proposed in this paper improves the accuracy by 0.9% and 1.49% respectively, and the parameters decrease by 47.6% and 34.02% respectively.

The result of SLIR in Tab.1 is better than SLIR-without-SA model without shuffle attention module, which shows that the group attention mechanism can improve the network. From the comparison of attention heat maps in Fig.11, it can be found that the SLIR model pays more attention to the key areas of pest pictures, which shows that the grouping attention mechanism not only improves the accuracy of the network but also enhances the interpretability of the neural network model, which makes the model pay more attention to the pests in the pictures, reduces the dependence on background information and reduces over-fitting.

4. Conclusion

In this paper, aiming at the problems of large amount of parameters and low recognition rate of traditional rice pest and disease recognition algorithm, a SLIR model is proposed by combining shuffle attention and Inception-ResNetV2 module. The results show that compared with the traditional algorithm, the SLIR model can achieve better recognition effect of rice diseases and insect pests with small parameters. Secondly, the results also show that the shuffle attention mechanism can optimize the convergence speed and accuracy of the network.

References

- (1) He K Zhang X Ren S et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- (2) Ballester P Araujo R. On the performance of GoogLeNet and AlexNet applied to sketches[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2016 30(1).
- (3) Szegedy C Ioffe S Vanhoucke V et al. Inception-v4 inception-resnet and the impact of residual connections on learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2017 31 (1).
- (4) Chen Y Li J Xiao H et al. Dual path net-works[J] . arXiv preprint arXiv:1707.01629 2017.
- (5) Sengupta A Ye Y Wang R et al. Going deeper in spiking neural networks: VGG and residual architectures[J]. Frontiers in neuroscience 2019 13: 95.
- (6) Xie S Girshick R Dollar P et al. Aggregated residual transformations for deep neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1492-1500.
- (7) Zhang H Wu C Zhang Z et al. Resnest: Split-attention networks[J]. arXiv preprint arXiv: 2004.08955 2020.
- (8) Woo S Park J Lee J Y et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-
- (9) Hu J Shen L Sun G. Squeeze-and-excitation net-works [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.
- (10) Wu W Zhang Y Wang D et al. SK-Net: Deep learning on point cloud via end-to-end discovery of spatial keypoints[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020 34(04): 6422-6429.

- (11) Lu Y Yi S Zeng N et al. Identification of rice diseases using deep convolutional neural networks[J]. Neurocomputing 2017 267: 378-384.
- (12) Liang W Zhang H Zhang G et al. Rice blast disease recognition using a deep convolutional neural network[J]. Scientific reports 2019 9(1): 1-10.
- (13) Zhang Q L Yang Y B. Sa-net: Shuffle attention for deep convolutional neural networks[C]//ICASSP 2021-2021 IEEE International Conference on Acoustic Speech and Signal Processing (ICASSP). IEEE2021: 2235-2239.