Fine-Grained Image Classification on Agricultural Pest Larvae

Zekai Cheng 1,a, Wan Xia 1,b

1 School of Computer Science and Technology, Anhui University of Technology, Anhui 243032, China
a chengzk@ahut.edu.cn, b wanx96@outlook.com

Abstract. Pest management is an essential part of the growth of crops. Accurately identifying the types of pests in the early stage is conducive to formulating targeted prevention and control measures to reduce pests' impact on grain production. In order to identify the pests in the larval stage as early as possible, in this paper, we compare the conventional classification model and the fine-grained classification model and construct a fine-grained image classification model that can be used to classify the larvae of crop pests, which improves the ability to identify pests in the larval stage. Experiments show that our optimized fine-grained classification model surpasses the general convolution image classification model on the fine-grained agricultural pest dataset AgrFIP20.

1. Introduction
China is a traditional agricultural country, and agricultural production occupies an essential position in the national economic production. Crop pests are one of the main factors affecting crop yield. Crops are eroded by different kinds of pests at all stages of growth, resulting in a decline in the final grain yield and quality. Therefore, the accurate identification of pest types in the early stage is beneficial to formulating reasonable control plans, improving the efficiency of killing pests, and protecting crops' growth.

Traditional agricultural pest identification often uses artificially designed features to match the target image texture, color, and posture. This method has great limitations and low efficiency and cannot meet the needs of modern agricultural production. After the popularization of machine learning methods, people try to use SVM, Naive Bayes, and other methods to classify and recognize pests, but these methods also need to preprocess the image to extract main features for classification. These methods also have certain shortcomings in the face of large amounts of data and complex scenarios. With the development of deep learning technology in recent years, various convolutional neural networks have been used for pest classification and recognition and have achieved good results. Compared with the previous methods, the method based on convolution neural networks eliminates the step of manually extracting features and can extract pests' features in complex environments. A well-trained model can classify faster and recognize more accurately than traditional methods.

At the current stage, the classification and recognition of pests using deep learning methods often cover different types with apparent differences. Such classification is also called conventional image classification, and fine-grained classification is to classify multiple sub-categories in the same category (the difference between the two see Fig. 1.). To kill pests as soon as possible, we need to identify pests in the larval stage accurately. Most of these larval pests appear in caterpillars, and
different pests have high similarities in larval morphology. Therefore, we apply the fine-grained classification to the larvae of fine-grained crop pests to improve larval pests' identification accuracy.

Fig. 1. Regular image classification (blue box) and fine-grained image classification (orange box).

Our main work: (1) Compare various existing general convolution image classification models and fine-grained image classification model on the fine-grained pest dataset. (2) Optimize the general CNN model and construct an image classification model suitable for fine-grained agricultural pests.

2. Related Works
With the development of deep learning technology in recent years, modern image classification technology in agricultural pests has also aroused people's interest. In 2017, Cheng et al. [1] introduced a residual error mechanism on Alexnet, which can classify and recognize pests in a complex farmland background, with a classification accuracy of 98.67%. Facing the specific pests that appeared in the grain storage environment, Cheng et al. [2] adopted a fine-tuning method, using Alexnext and GoogleNet with pre-training parameters to identify seven specific pests, and all reached 97.61% accuracy. Thenmozhi et al. [3] used the transfer learning method to identify pests on the public agricultural insect datasets NBAIR, Xie1, and Xie2, and the comprehensive highest accuracy reached 96.75%, 97.47%, and 95.97%, respectively. Chen et al. [4] improved the residual network and added the Bayesian method. Compared with the original model, the accuracy rate was improved by 9.6% on the same garden pest dataset. Aiming at dense aphids, Li et al. [5] proposed a coarse-to-fine network with a two-stage method, using a convolutional network to locate roughly dense areas and then using another convolutional network to identify pests. In 2020, Li et al. [6] also adopted a fine-tuning method to use GoogleNet on a manually collected pest dataset to identify pests in complex backgrounds, which achieved a performance improvement of 6.22% compared with the latest model at the time. Nanni et al. [7] combined the saliency method with the convolutional neural network model, and the classification accuracy reached 61.93% on the large public dataset IP102 [8] containing 102 categories.

At present, fine-grained image classification has vast applications in some actual scenes that require detailed classification, so it is very suitable for the classification of pest larvae. It is a classic fine-grained classification method to extract the corresponding local features and global features based on the target part. Berg et al. [9] obtained the features of different parts for classification with manually marked parts. Some methods in [10-12] found the best part and performed more accurate semantic segmentation and feature fusion to obtain better fine-grained part feature representation. Some scholars [13-15] used weakly supervised and unsupervised methods to extract part features, which reduced the cost of human annotation. Some methods combining the attention mechanism [16-17] have also been proposed to extract more relevant part features through the attention mechanism. Unlike the method that focuses on parts, the fine-grained classification method based on end-to-end feature coding focuses on extracting high-level representations and interaction relationships of object features. Lin et al. [18] used the bilinear CNN model for fine-grained classification and achieved good classification results. After that, researchers [19-21] made a series of improvements to reduce the computational complexity of bilinear operations. Besides, Sun et al. [22] proposed a random peak suppression method, which forced the network to find more distinct areas by randomly suppressing some areas of the feature map, which improved fine-grained object recognition accuracy.
From the development of research in the above two fields, it can be seen that the application of conventional deep learning image classification technology in agricultural pests has become more and more common. However, most of them classify pests in the adult stage, and the classification difficulty is also relatively low. We can classify these similar larval pests more accurately with fine-grained classification methods than general image classification models.

3. Methods

In this section, we introduce the models and methods used in the experiment. The classic and excellent ResNet18 [23] is suitable for conventional image classification. BLCNN [18] and the random peak suppression method proposed by Sun et al. [22] are designed for fine-grained image classification.

3.1. ResNet18

To solve the problem that the gradient of the deep model becomes smaller and smaller in the later training, He et al. [23] proposed the ResNet series (the network structure of ResNet18 is shown in Fig. 2.a) model, which uses residual connections to avoid network degradation. The structure of the residual connection is shown in Fig. 2.b.

![Fig. 2. ResNet18 (a) and residual connection (b).](image)

3.2. BLCNN

Compared with other fine-grained algorithms that require part labeling, BLCNN [18] is based on end-to-end training, without pre-labeling the samples, and can directly use the original pictures to train the model. In terms of feature extraction, BLCNN uses two CNN streams to extract features in two branches, and finally utilizes the outer product to model the local paired features of the image, and obtains the fusion feature containing the interaction relationship of different channels, which improves the model’s ability to learn fine-grained level object textures. BLCNN was ahead of other algorithms in the standard fine-grained dataset at the time.

![Fig. 3. BLCNN model structure.](image)

3.3. Random peak suppression

Also, in the field of fine-grained classification, Sun et al. [22] proposed random peak suppression and block suppression on the feature map, forcing the network to find other potential areas to obtain more subtle differences for classification. In Fig. 4, the peak is located at the lower center of the original feature map on the left. Except for the feature map's peak area, the other three areas are also randomly selected and suppressed. In [22], the author used the Bernoulli binomial distribution to randomly select the areas that need to be suppressed. The values of all selected areas in the feature map will be
multiplied by the suppression factor $\alpha$ to reduce its contribution to the overall feature map. In their paper, $\alpha = 0.1$.

![Feature map after random peak suppression and block suppression](image)

Fig. 4. Feature map after random peak suppression and block suppression (the original image is added below the feature map for comparison with the original image).

4. Experiments

4.1. Data augmentation

Our dataset comes from the Anhui Academy of Agricultural Sciences. The fine-grained pest dataset $AgrFIP20$ is a subset of the large-scale agricultural pest dataset $ArgIP138$, covering 20 categories and including 1984 training pictures and 512 testing pictures. $AgrFIP20$ only contains the larvae of various pests, namely caterpillars, which can meet the needs of fine-grained image classification.

To allow the model to learn features from more angles to improve the model's generalization ability, we use data augmentation during training and perform random stretching, flipping, rotating, and zooming operations on the original pest samples. Fig. 5. shows the data augmentation.

![Data augmentation](image)

Fig. 5. Data augmentation ($AgrFIP20$).

The upper part is the original picture, and the lower part is the enhanced picture.

4.2. Implementation details

We rebuild all experimental models under the framework of Tensorflow 2.3.0. The input size of each model is 224×224, and all models are trained from scratch. The Adam optimizer's initial learning rate is 0.001, and the subsequent learning rate adopts the cosine decay schedule. Each experiment was performed on a Tesla V100 GPU separately.

4.3. Results

4.3.1. Result analysis We experimented with five different models on the fine-grained agricultural pest dataset $AgrFIP20$, namely ResNet18 and ResNet34 of the ResNet series [23], and the optimization model ResNet18_RPS and ResNet34_RPS that we built with the addition of the peak suppression method [22], and the classic algorithm BLCNN [18] on fine-grained classification. The detailed experimental results are shown in Table 1.

| Methods  | Backbone | Parameters | Top-1 accuracy (%) |
|----------|----------|------------|-------------------|
| BLCNN    | VGG16    | 19.9M      | 82.81             |
| ResNet18 | Resnet18 | 11.2M      | 89.84             |

Table 1. Experimental results of five models on $AgrFIP20$. 


From the results in Table 1, it can be seen that thanks to the advantages of residual connection [23], the performance of the model using ResNet as the backbone is higher than the model using VGG16 as the backbone. Even if BLCNN [18] is optimized for fine-grained classification, the classification accuracy (82.81%) is still not as good as the general image classification models ResNet18 (89.84%) and ResNet34 (89.65%). Compared with the original ResNet model, ResNet18_RPS (94.14%) and ResNet34_RPS (89.84%) we built with peak suppression methods increased by 4.3% and 0.19%, respectively. From the parameters of these five models, it can also be found that higher performance is not obtained with more layers and more parameters. For AgrFIP20 with only 20 classes, the recognition accuracies of shallower Resnet18 and ResNet18_RPS are higher.

4.3.2. Class activation map analysis Class activation diagram is a way to visually show which area on the sample the model pays attention to. To analyze the ability of various models to recognize pests in different environments, Fig. 6. shows how the three types of models pay attention to larvae in five different scenarios.

4.3.3. Training curve analysis Fig. 7. shows the comparison of the training curves of our fine-grained model and other models on the dataset AgrFIP20. First of all, we can see that the training accuracy and validation accuracy of BLCNN are lower than other models. It can also be seen from the training curves represented by the solid lines that the upward trend of the training curves of ResNet18_RPS and ResNet34_RPS with the peak suppression method added are lower than those of the general ResNet model, which proves that the peak suppression method alleviates overfitting. At the end of the dotted lines, the accuracy of the validation of ResNet18_RPS is ultimately higher than those of other models, which conforms to the experimental results in Table 1, which proves that the optimized model we constructed can effectively recognize pest larvae.
5. Conclusion

Pests with caterpillar morphology often look similar and are difficult to distinguish. To accurately identify these pests in the larval stage, we combine fine-grained classification methods and build an agricultural fine-grained image classification model to classify pest larvae. We compare the general image classification model, the existing fine-grained classification model, and the model we built on the fine-grained agricultural pest dataset, AgrFIP20, composed of pest larvae. Experiments show that the model constructed in this paper has a higher recognition accuracy on AgrFIP20 than the general image classification model and existing fine-grained classification model.

References

[1] Cheng, X., Zhang, Y., Chen, Y., Wu, Y., & Yue, Y. (2017). Pest identification via deep residual learning in complex background. Computers and Electronics in Agriculture, 141, 351-356.
[2] Cheng, X., Wu, Y. Z., Zhang, Y. H., & Yue, Y. (2018). Image Recognition of Stored Grain Pests: Based on Deep Convolutional Neural Network. Chinese Agricultural Science Bulletin, 34(1), 154-158.
[3] Thenmozhi, K., & Reddy, U. S. (2019). Crop pest classification based on deep convolutional neural network and transfer learning. Computers and Electronics in Agriculture, 164, 104906.
[4] Chen, J., Chen, L. Y., Wang, S. S., Zhao, H. Y., & Wen, C. J. (2019). Pest Image Recognition of Garden Based on Improved Residual Network. Transactions of the Chinese Society for Agricultural Machinery, 5.
[5] Li, R., Wang, R., Xie, C., Liu, L., Zhang, J., Wang, F., & Liu, W. (2019). A coarse-to-fine network for aphid recognition and detection in the field. Biosystems Engineering, 187, 39-52.
[6] Li, Y., Wang, H., Dang, L. M., Sadeghi-Niaraki, A., & Moon, H. (2020). Crop pest recognition in natural scenes using convolutional neural networks. Computers and Electronics in Agriculture, 169, 105174.
[7] Nanni, L., Maguolo, G., & Pancino, F. (2020). Insect pest image detection and recognition based on bio-inspired methods. Ecological Informatics, 57, 101089.
[8] Wu, X., Zhan, C., Lai, Y. K., Cheng, M. M., & Yang, J. (2019). Ip102: A large-scale benchmark dataset for insect pest recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 8787-8796).
[9] Berg, T., & Belhumeur, P. N. (2013). Poof: Part-based one-vs.-one features for fine-grained categorization, face verification, and attribute estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 955-962).

[10] Zhang, N., Donahue, J., Girshick, R., & Darrell, T. (2014, September). Part-based R-CNNs for fine-grained category detection. In European conference on computer vision (pp. 834-849). Springer, Cham.

[11] Huang, S., Xu, Z., Tao, D., & Zhang, Y. (2016). Part-stacked cnn for fine-grained visual categorization. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1173-1182).

[12] Wei, X. S., Xie, C. W., Wu, J., & Shen, C. (2018). Mask-CNN: Localizing parts and selecting descriptors for fine-grained bird species categorization. Pattern Recognition, 76, 704-714.

[13] Simon, M., & Rodner, E. (2015). Neural activation constellations: Unsupervised part model discovery with convolutional networks. In Proceedings of the IEEE international conference on computer vision (pp. 1143-1151).

[14] He, X., & Peng, Y. (2017, February). Weakly supervised learning of part selection model with spatial constraints for fine-grained image classification. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 31, No. 1).

[15] Wang, Z., Wang, S., Li, H., Dou, Z., & Li, J. (2020, April). Graph-Propagation Based Correlation Learning for Weakly Supervised Fine-Grained Image Classification. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 07, pp. 12289-12296).

[16] Zheng, H., Fu, J., Mei, T., & Luo, J. (2017). Learning multi-attention convolutional neural network for fine-grained image recognition. In Proceedings of the IEEE international conference on computer vision (pp. 5209-5217).

[17] Zheng, H., Fu, J., Zha, Z. J., & Luo, J. (2019). Looking for the devil in the details: Learning trilinear attention sampling network for fine-grained image recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 5012-5021).

[18] Gao, Y., Beijbom, O., Zhang, N., & Darrell, T. (2016). Compact bilinear pooling. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 317-326).

[19] Cui, Y., Zhou, F., Wang, J., Liu, X., Lin, Y., & Belongie, S. (2017). Kernel pooling for convolutional neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2921-2930).

[20] Kong, S., & Fowlkes, C. (2017). Low-rank bilinear pooling for fine-grained classification. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 365-374).

[21] Zheng, H., Fu, J., Zha, Z. J., & Luo, J. (2019). Learning deep bilinear transformation for fine-grained image representation. arXiv preprint arXiv:1911.03621.

[22] Sun, G., Cholakkal, H., Khan, S., Khan, F., & Shao, L. (2020, April). Fine-grained recognition: Accounting for subtle differences between similar classes. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 07, pp. 12047-12054).

[23] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).