Recognition of Common Insect in Field Based on Deep Learning

Xu Cao, Ziyi Wei, Yijie Gao, Yingqiu Huo*

College of Information Engineering, Northwest A&F University, Yangling, Shaanxi, 712100, China

*Corresponding author e-mail: fallying@nwsuaf.edu.cn

Abstract. Traditional insect taxonomy methods have high technical requirements, and the ability of artificial identification of insects is insufficient. In order to solve the problem, this paper proposes a method of common insects recognition in field based on transfer learning. A total of 9 kinds of insects, such as mythimna separata, rice borer, rice plant hopper, mole cricket, mantis, locust, grass fly, ladybug, and ditch beetle are collected for classification and identification, those include the main insect pests and some natural enemies of the main food crops in the field, such as wheat, rice, corn, etc. Then we use the digital image processing technology and the confrontation generation network to expand the insect dataset, and build a model based on transfer learning to transfer the knowledge learned by VGG16, VGG19, InceptionV3, and InceptionV4 on the ImageNet dataset to the insect classification and recognition. Experimental results show that the transfer learning training model has better classification performance and higher convergence speed, and data expansion can help extend sample and avoid overfitting. The highest recognition accuracy is up to 97.39% among models, which adopt the VGG19 convolutional neural network to pretrain the model for transfer learning. This method has the high recognition accuracy, less time consumption, simple and convenient, robustness in particular for the translation and rotation, which provides a reference for the identification and classification method of field insects.

1. Introduction

Identification of plant diseases and insect pests is a basis for plant protection and utilization. In the process of production and planting, farmers correctly identify the occurrence of plant diseases and insect pests, physiological diseases, natural enemies, and grasp the habits and characteristics of pests and natural enemies, so as to avoid the problems of weak plant protection knowledge and blind application of pesticides. Agricultural pest detection and identification is an important part of precision agriculture and modern agriculture. Traditional pest detection depends on the experience of the staff. It has a large workload, poor timeliness, and high knowledge reserve requirements, which cannot meet the demand of pest identification and detection[1-3]. Therefore, it is of great significance to explore the principles and methods of automatic identification of insects by computers.

In recent years, with the rapid development of computer vision technology, it has provided a technical basis for the realization of automatic insect recognition. Among them, research on automatic insect recognition based on image processing technology has become a hot spot[4-7]. The main techniques of insect classification currently used in production are as follows[8-13]:1. Morphological classification method, using electron microscope technology to observe and analyze insect morphology; 2. Behavioral classification method to identify the types of insects by analyzing some
behavioral characteristics of insects. 3. Genetic classification method, through the comparative analysis of insect chromosome number and morphology, chromosome grouping type, sex chromosome position, total heterochromatin content and chromosome meiosis behavior patterns, to identify the type of insect; 4. Biochemical classification method, using isozyme electrophoresis and epidermal hydrocarbons for insect classification; 5. Molecular biology methods, using nucleic acid sequence analysis technology, molecular hybridization technology, RFLP method and other technologies to classify and identify insects; 6. Use computer calculation to classify insects. Both the traditional morphological observation method and the current emerging classification method combining multiple biotechnologies require the classifier to possess rich professional knowledge and other theoretical methods related to classification, which have high technical requirements. Computer technology has achieved certain achievements in insect classification and recognition. Scholars at home and abroad use BP network to realize automatic insect recognition[14-18], and achieved satisfactory results. Cai Xiaona[19], Pan Pengliang[20], etc. use the features of the upper wings of insect images to classify insects such as Noctuidae and butterflies. Image-based pest recognition objects are mostly lepidopteran insects. However, the period of pest damage to the main food crops in the field is mostly the larval period. Due to the vast territory of my country, the large number of agricultural production activities, the strong mechanical nature and the rapid cycle change, a small number of classification experts and high-demand identification environments cannot meet the needs of a large number of cumbersome classification identification work.

To solve the above problems, this paper proposes a method for identifying common insects in the field based on transfer learning. And the digital image processing technology and confrontation neural network method are adopted to solve the problem of insufficient insect sample. By comparing the results of different training parameters of multiple convolutional models, the model with the best recognition result and the corresponding parameters are explored. A controlled experiment method was used to obtain data using the horizontal and vertical experimental methods. This paper used the TensorFlow framework to achieve model training, model testing and model result output.

2. Data collection and processing

2.1. Collection of images

In this study, the main insect pests and their natural enemies of the main food crops in the field (wheat, rice, corn, etc.) are selected. The main insect pests and some of their natural enemies of the above food crops are selected from 9 types of insects (mythimna separata, rice borer, rice planthopper, mole cricket, mantis, locust, grass fly, ladybug, ditch beetle) as the training objects of the project. Insect images were collected using a combination of web crawlers and laboratory photography. According to the needs of this experiment, comparative analysis is carried out according to the crop disease and insect pest map, and the images obtained by integrating the two image acquisition methods are used to filter out the insect pictures: There were 436 locusts, 384 grasshoppers, 314 mantises, 307 ladybirds, 139 mole crickets, 114 rice planthoppers, 110 mythimna separata, 83 furrow needles, and 80 rice borers. In view of the changeable growth patterns of insects, we choose the damage period as the research object morphology, as shown in Figure 1.
2.2. Data augmentation

2.2.1. Digital image processing. First copy 10 copies of each image with a uniform size of 256*256 pixels, and then perform the following affine transformation on the image. Each image is randomly selected from 0-5 image processing methods in the following 10 methods, as shown in Figure 2:(1) Flip and transform the image; (2) Randomly use one of Gaussian blur, mean blur and median blur to enhance the picture; (3) Sharpening; (4) Edge detection, the detected assignment 0 or 255 and then superimposed on the original image; (5) Noise disturbance, add Gaussian noise to the picture; (6) Rotation transformation/ reflection transformation; (7) Contrast transformation to change the contrast of the entire image to half or double the original; (8) Change RGB into grayscale image and multiply it by alpha to add to the original image; (9) Move the pixels to the surrounding area; (10) Distort local areas of the image. Part of the image after processing is shown in Figure 3. Using the digital image processing method to simulate the same image with different angles and different backgrounds can greatly increase the scale of the data set and improve the classification recognition rate.
2.2.2. Generative confrontation network. Due to the small amount of original data, the number of data sets amplified by traditional digital image processing methods is limited. Excessive amplification will lead to overfitting of training results. To solve this problem, we try to use DCGAN to generate insect images for augmenting the data set.

DCGAN adds a deep convolutional network structure on the basis of GAN to specifically generate image samples[21], as shown in Figure 4.

![DCGAN structure diagram](image)

Figure 4. DCGAN structure diagram

Assuming that the real picture data used for training is \( x \), The real data distribution learned by \( G \) is \( P_{\text{data}}(x) \). The distribution of noise \( z \) is set as \( P_z(z) \). According to the cross picking loss, the following loss function can be constructed as formula (1)

\[
V(D, G) = E_{x \sim P_{\text{data}}(x)}[\ln D(x)] + E_{z \sim P_{z}(z)}[\ln(1 - D(G(z)))]
\]

(1)

In actual training, the gradient descent method is used to optimize \( D \) and \( G \) alternately. The detailed steps are as follows:

Step 1: Select some samples \( \{z^{(1)}, z^{(2)}, \cdots, z^{(n)}\} \) from the known noise distribution \( P_z(z) \)

Step 2: Select the same number of real pictures \( \{x^{(1)}, x^{(2)}, \cdots, x^{(n)}\} \) from the training data.

Step 3: Set the parameter of discriminator \( D \) to \( \theta_{D} \), find the gradient of the loss parameter as formula (2), and add this gradient when \( \theta_{D} \) is updated.

\[
\nabla \frac{1}{n} \sum_{i=1}^{n} [\ln D(x^{(i)}) + \ln(1 - D(G(z^{(i)})))]
\]

(2)

Step 4: Set the parameter of generator \( G \) to \( \theta_{g} \), find the gradient of the loss parameter as formula (3), and subtract the gradient when updating \( \theta_{g} \).

\[
\nabla \frac{1}{n} \sum_{i=1}^{n} [\ln(1 - D(G(z^{(i)})))]
\]

(3)

Since \( D \) is to hope that the greater the loss, the better, \( G \) is to hope that the smaller the loss, the better, so one of them is to add a gradient, another is to subtract the gradient. When the training is completed, a noise can be randomly taken from \( P_z(z) \), and a new sample conforming to \( P_{\text{data}}(x) \) can be generated after the \( G \) operation.

Part of the pictures amplify by the DCGAN method are shown in Figure 5. DCGAN is played by the generator and the discriminator, so that the pictures generated by the generator are enough to "fake the truth" to achieve the purpose of augmenting the data set.
3. Experimental process and analysis

3.1. Design of experimental

Under the premise of ensuring multiple samples, we look for multiple network models for sample training. We load each pre-model into the network and train full-layer parameters and load the network to train only output layer parameters. Then, various types of models are trained with sample rates of $1 \times 10^{-3}$, $5 \times 10^{-4}$, and $1 \times 10^{-4}$, and the results obtained by training models with the same layer parameters are compared. The best class among the training full-level parameter models is selected, and only the outer-level parameter models are trained in the same way. Amplify the training rounds of the two optimal models, control other variables unchanged, conduct training, determine the convergence curve, obtain the convergence information, and ensure that the non-amplified round models have all converged. After obtaining a model with an increased number of training rounds, compare only the training outer-layer parameters and the training full-layer parameter models to obtain the difference in accuracy and training time, and analyze the results.

3.1.1. Model selection and analysis. The VGG model performs well in transfer learning tasks. To extract features from images, the VGG model is the preferred algorithm. The VGG model has been improved on the basis of AlexNet, using a deeper network structure. The entire network uses a 3*3 convolution kernel size and a 2*2 maximum pooling size. It makes it possible to control the number of parameters while obtaining more image features, avoiding excessive calculations and overly complicated structures. Because the VGG model uses the depth of the network structure to obtain a better recognition rate, this experiment selects the deeper network structure D and E in Figure 6, namely the VGG16 and VGG19 models.

![Figure 5. DCGAN generated graph](image)

![Figure 6. VGG network structure diagram](image)
The VGG model obtains high-quality effects by increasing the model depth. This will cause some problems, such as too many parameters and too much calculation complexity, which is difficult to apply; the deeper the network, the easier the gradient disappears, and it is difficult to optimize the model. In order to maintain the sparseness of the network structure and take advantage of the high computational performance of the dense matrix, GoogleNet uses the inception module as a unit, as shown in Figure 7, using different size convolution kernels to obtain different sizes of receptive fields, and finally in FilterConcat The stitching in the channel dimension means the fusion of features of different scales. This time choose Inception_v3 and Inception_v4 to experiment.

Figure 7. Inception module

3.1.2. Training method. In order to explore the effect of each model on field identification of insects to obtain a better model, four pre-models of VGG16, VGG19, InceptionV3, and InceptionV4 are now used for training. And fine-tune the learning rate:

- The initial learning rate is set to $1 \times 10^{-3}$. After the result is obtained, the learning rate is lowered to $5 \times 10^{-4}$ for training. If the accuracy rate increases after the learning rate is lowered, the learning rate is again reduced to $1 \times 10^{-4}$. It drops three times in a row, and if it finds that the accuracy rate drops, it finds a local optimal solution.

- If the learning rate is reduced three times and the accuracy rate has not been found to be reduced, the learning rate obtained for the third time is used as the local optimal solution.

- If the accuracy rate drops when the learning rate is lowered for the first time, the learning rate is increased. If the accuracy rate increases after the adjustment, the learning rate is increased three times in a row. Until a certain learning rate decreases, a local optimal solution is found.

- If the local optimal solution is not found if the learning rate is increased three times, the third increased learning rate is used as a substitute.

- If the accuracy decreases when the learning rate is increased for the first time, the initial learning rate is the local optimal solution.

The local optimal solution and the two left and right learning rates are used as variables in this experiment, and InceptionV3 only trains the outer parameters as the object model for exploring the learning rate.

Each pre-model is divided into two cases of loading and training the full-layer parameters of the pre-model and loading the pre-model full-layer parameters and training only the outermost parameters of the pre-model. Analyzing the differences and trade-offs between time and accuracy in the two cases helps to explore whether to sacrifice part of the time or part of the accuracy for the purpose in a
particular situation. A total of 24 network models are obtained. The static learning rate is adopted, the rmsprop optimization algorithm is selected, the weight attenuation is set to $4 \times 10^{-5}$, the batch size is 32, and 10000 rounds of training are performed. Based on the results of 24 models, analyze the trend of result changes due to parameter changes. And by comparing and selecting the two types of models with the best performance in the outer and full layer parameter training, the training cycle is expanded by 100,000, and the other variables are controlled unchanged to train, and again 6 network models are obtained. They are used to clearly observe the convergence of the two optimal models, and record and analyze the training completion time of different models.

3.2. Results and analysis

Four models are trained to obtain 24 model results. As shown in Table 1, the classification effects of field insects in the four types of models are sequentially decreased by VGG19, InceptionV4, InceptionV3, and VGG16, and the highest accuracy rate of VGG19 is 97.3913%. In the same type of model, the model classification effect obtained by loading the network and training all parameters is significantly better than the effect of training only the outer parameters. When the other variables are controlled unchanged, the accuracy difference can be up to 11.82726%. From the trend of fine-tuning the learning rate, the lower the learning rate, the higher the accuracy.

| ID | Pre-trained model | Training level | Learning rate | Training loss | Verification accuracy |
|----|------------------|----------------|---------------|---------------|-----------------------|
| 1  |                  |                | 0.0001        | 0.3779        | 96.0435%             |
| 2  |                  |                | 0.0005        | 0.3514        | 93.4348%             |
| 3  | Inception_V3     | Full           | 0.001         | 0.2617        | 94.6957%             |
| 4  |                  | Logits,        | 0.0001        | 1.1629        | 84.2609%             |
| 5  |                  | AuxLogits      | 0.0005        | 0.5905        | 88.2174%             |
| 6  |                  |                | 0.001         | 0.4064        | 83.3478%             |
| 7  |                  |                | 0.0001        | 0.6137        | 96.0435%             |
| 8  | Inception_V4     | Full           | 0.0005        | 0.5443        | 94.7391%             |
| 9  |                  |                | 0.001         | 0.4342        | 94.7391%             |
| 10 |                  | Logits,        | 0.0001        | 1.1771        | 88.2174%             |
| 11 |                  | AuxLogits      | 0.0005        | 0.7304        | 90.8261%             |
| 12 |                  |                | 0.001         | 0.7935        | 92.1304%             |
| 13 | VGG19            | Full           | 0.0005        | 0.0548        | 97.3913%             |
| 14 |                  |                | 0.001         | 0.0884        | 92.1304%             |
| 15 |                  | fc8            | 0.0005        | 0.4422        | 92.1304%             |
| 16 |                  |                | 0.0001        | 0.2701        | 93.4348%             |
| 17 |                  | fc8            | 0.0005        | 0.5017        | 96.0870%             |
| 18 | VGG16            | Full           | 0.0005        | 0.0742        | 88.2171%             |
| 19 |                  |                | 0.001         | 0.0758        | 90.8261%             |
| 20 |                  |                | 0.0001        | 0.2571        | 92.1304%             |
| 21 |                  | fc8            | 0.0005        | 0.2657        | 92.1304%             |
| 22 |                  |                | 0.001         | 0.376         | 90.8261%             |
| 23 |                  |                | 0.001         | 0.376         | 90.8261%             |
| 24 |                  |                | 0.001         | 0.376         | 90.8261%             |
Taking InceptionV3 as an example, the top three of the six lines are the model loss of the training output layer only, and the bottom three are the model loss of the training full layer. The learning rate of the three lines of each part from top to bottom is $1 \times 10^{-3}$, $5 \times 10^{-4}$, $1 \times 10^{-4}$. It can be found from Figure 8 that the higher the learning rate, the more obvious the loss decline and the faster the convergence. However, when the models are all in convergence, the lower the learning rate, the smaller the training loss, and the higher the accuracy. The model VGG19 with the best recognition effect is expanded ten times the number of training batches, keeping the learning rate unchanged. The results obtained are shown in Table 2. When the number of training rounds is 100,000 and the number of training samples in the data set is 13842, the accuracy of VGG19 differs by about 3.9%, and the time difference is about 9 hours.

### Table 2. Results of batch deepening model

| ID | Pre-trained model | Training level | Learning rate | Training loss | Verification accuracy | Training time |
|----|-------------------|----------------|---------------|---------------|-----------------------|---------------|
| 1  | VGG19             | full           | 0.001         | 0.0505        | 0.8947826             | 12:21:50      |
| 2  | VGG19             | full           | 0.0005        | 0.0498        | 0.90826088            | 12:41:20      |
| 3  | VGG19             | full           | 0.0001        | 0.0498        | 0.92130435            | 12:20:50      |
| 4  | VGG19             | fc8            | 0.001         | 0.1258        | 0.93434781            | 3:54:10       |
| 5  | VGG19             | fc8            | 0.0005        | 0.2211        | 0.94739133            | 3:37:40       |
| 6  | VGG19             | fc8            | 0.0001        | 0.5960        | 0.94739133            | 3:40:40       |

After controlling other variables, the training losses of the training batches of 5000, 10000, 50000, and 100,000 are shown in Figure 9. It can be seen that the curve has begun to converge when the batch is 10000, and there is no case where the training batch is too low.

### 4. Conclusion

In this study, nine common field insects such as the ladybug, locust, mantis, and the lacewing insect are collected. The combination of traditional image processing methods and the generation of adversarial networks are adopted to extend the sample data set, which solve the problem of insufficient insect sample data set. Among the four models for horizontal comparison in the preliminary identification, the accuracy of vgg19 model with the best effect reached 97.3913%. A longitudinal comparison of a single model shows that with the same number of training rounds, the higher the learning rate, the faster the training loss drops, but after the model converges, the effect is
slightly lower than the model with a lower learning rate. The result of performing full-layer parameter training on the pre-model is always better than the result of training only the output layer, and the accuracy difference of vgg19 is about 3.9%. However, under the experimental conditions in this paper, the training time of the full layer parameters is about 4 times that of the outer layer training only. Therefore, the full-layer or outer-layer training method can be selected according to different training needs.

Acknowledgments
This article is supported by the National Scientific and Technological Innovation Program for College student in 2019 (No. 2019071209), National Key Research and Development Project (2017YFC0403203), Open subject of the Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs (2018AIOT-09), Chinese National Natural Science Foundation grant number 41771315, and Key Research and Development program of Shaanxi Province (2020NY-098)

References
[1] Liu Haiqi. Drive agricultural modernization with precision agriculture to accelerate the digital transformation of modern agriculture [J]. Agricultural Resources and Regional Planning in China, 2019, 40(1): 1-6, 73.
[2] Nie Juan, Sun Ruizhi, Deng Xuefeng, Yang Hao. Accurate Agriculture Uncertainty Complex Event Processing Based on Data Lineage Management [J]. Journal of Agricultural Machinery, 2016, 47(5): 245-253.
[3] Feng Hongqiang, Yao Qing. Automatic identification and monitoring technology of agricultural pests [J]. Plant Protection, 2018, 44(5): 127-133,198.
[4] Chen Xiaolin, Hou Xinwen, Liu Chenglin, etc., 2008 Insect Image Automatic Identification Technology.[J]. Insect Knowledge, 45(2): 315–316.
[5] Xian Dingxiang, Yao Qing, Yang Baojun, Luo Ju, Tan Chang, Zhang Chao, Xu Yicheng. Research on automatic recognition technology of rice lamp trap insects based on image[J]. China Rice Science. 2015(03).
[6] Yao Kan, Xu Peng, Zhang Guangqun, Wang Hangjun, Review of image-based insect classification and recognition research.[J] Intelligent Computer and Application.2019(03).
[7] Karunakaran C, Jayas D S,White A N D G.Identification of Wheat Kernels damaged by the Red Flour Beetle using X-ray Images.Biosystems Engineering,2004.
[8] Zha Yuping, Luo Qigui. Application of modern technology in insect classification[J]. Jiangxi Forestry Science and Technology, 2005, (1):34-36.
[9] Wen Shuoyang, Peng Tongxu, Xie Li, et al. Study on the reproductive isolation of Lin's fruit fly[J]. Zoological Research, 1997, 18 (1): 99-103.
[10] Li Zhiqin, Yan Hongbo, Li Chengde.Main technical means of insect classification[J].Research on Hebei Forest and Fruit,2006(04):398-403.
[11] Li Shaowen, Meng Yuping, Li Jubei. A comparative study on esterase isozymes of Hymenoptera insects[J].Acta Entomologica Sinica, 1987, 30 (3) :266-269.
[12] Chapco W.Mitochondrial DNA variation in North America mekanoplinegrasshoppers[J].Heredity, 1994, 72 (2) :1-9.
[13] Wang Jian, Zhou Guomin. Research on the identification system of water borer borer based on neural network. Agricultural Network Information, 2006.
[14] Wang Yinglong, Dai Xiangliang. Application of image processing technology in rice pest system. Microcomputer Information, 2007.
[15] Yang Lingling. Research on Automatic Identification Technology of Rice Planthopper [D]. Nanjing: Nanjing Agricultural University, 008.
[16] Singh C B, Jayas D s, Paliwal J, White N D G. Identification of insect-damaged wheat kernels using short-wave near-infrared hyperspectral and digital colour imaging. Computers and Electronics in Agriculture, 2010.

[17] Cai Qing, He Dongjian. Vegetable leaf-eating pest recognition technology based on image analysis. Computer Application, 2010.

[18] Zhen Tong, Fan Yanfeng. Research on classification and recognition technology of stored grain pests based on support vector machine. Computer Engineering, 2006.

[19] Cai Xiaona, Huang Dazhuang, Shen Zorui, et al. Application of the characteristics of moth wings and veins in the mathematical taxonomy of Spodoptera litura[J]. Journal of China Agricultural University, 2015, 20 (5) :188-193.

[20] Pan Pengliang, Yang Hongzhen, Shen Zorui, et al. Application of mathematical morphological characteristics of wing veins in butterfly classification and identification[J]. Acta Insect Taxonomy, 2008, 30 (2) :151-160.

[21] He Zhiyuan, (2018) 21 projects play deep learning. Electronic Industry Press, Beijing