Research on Intelligent Detection Method of Crop Diseases and Pests based on Machine Learning

Heng Xiao
Sanya University, Sanya, 572022, China

Abstract: In order to solve the problem of poor detection effect of traditional crop diseases and insect pests, an intelligent detection method of crop diseases and insect pests based on machine learning was proposed. Based on the principle of machine learning, the characteristics of common plant diseases and insect pests are collected and analyzed, and the detection algorithm is optimized by combining the migration learning method, so as to effectively detect the damage of plant diseases and insect pests. Finally, experiments prove that the intelligent detection method of crop diseases and insect pests based on machine learning has higher accuracy than the traditional method.

Keywords: Intelligent decision-making; Fuzzy neural network; Area detection; Precision agriculture

1. Introduction

Agricultural production is a vital part of China's economic development. Its output will not only affect the country's economic development, but also affect the quality of people's daily life. However, because the traditional detection of crop diseases and insect pests relies too much on experts' experience, low efficiency and lack of intelligence are important limitations that affect the detection performance of crop diseases and insect pests. With the continuous development of computer vision and neural network algorithms, and with the continuous popularization of smart phones, the intelligent detection of crop diseases and insect pests has attracted more and more attention and has been developed rapidly. With the rapid development of current science and technology, the current machine learning technology is closely related to the rapid development of modern agriculture [1]. Therefore, a detection method of crop diseases and insect pests based on machine learning is proposed, which has good effect on classification and identification of crop diseases and insect pests. Firstly, the detection method based on machine learning is introduced. This part mainly elaborates the structural design of machine learning and the training details of machine learning, and lists the classification of crop disease and insect pests data set by the detection method based only on machine learning, and analyzes the causes of over-fitting. Then the improvement of detection method based on transfer learning idea is introduced. It can be seen that the idea of transfer learning can alleviate the over-fitting problem to a large extent and greatly improve the accuracy of classification and detection. After that, the support vector machine is used to replace the original softmax classification layer for classification. it can be seen that this method can improve the accuracy rate slightly and has certain effect on alleviating the over-fitting problem [2]. Finally, the significance of data expansion for alleviating over-fitting problem and improving the detection accuracy of diseases and insect pests is introduced, and relevant test conclusions are given. In addition, this paper also applies the proposed detection method of crop diseases and insect pests to the data set established by ourselves, which can also achieve higher classification accuracy, indicating that the method has better robustness. Compared with traditional methods, the detection method of crop diseases and insect pests based on machine learning in visible light images has good detection performance for diseases and insect pests detection, and its intelligence and high accuracy have good effects for solving practical problems.

2. Intelligent Detection Method for Crop Diseases and Insect Pests

2.1. Collection of characteristics of crop diseases and insect pests

In the process of intelligent detection of crop diseases and insect pests, it is necessary to collect and analyze the image feature samples of common crop diseases and insect pests. During the research, most of the crop information was collected from Plant Village website. The main function of this website is to provide diagnostic services for crop diseases and insect pests for the number of pictures of farmers around the world [3]. Plant Village has an official, open source and free image sample database, which contains dozens of plants, each of which contains several states of health and certain diseases, totaling nearly 100,000 visible light pest images. The image data set used is selected from this open source database. Since only a small number of training samples can be obtained in practice, 37 categories are selected here, with a maximum of 200 pictures for each category, and a total of 7,062 pictures are used as image data sets of crop diseases and insect pests. In the process of sample collection, the more complex the network structure is, the more feature maps are obtained. When the task of identification is very complex, higher classification accuracy can often be achieved. However, the principle of machine learning is not as complex as possible [4]. If the number of samples in machine learning is too large, the more information about image characteristics will be extracted, resulting in an increase in dimensions, a greater burden on the calculation, and a more serious overfitting problem. Therefore, the samples of crop diseases and insect pests are analyzed based on the principle of machine learning. The specific information of the data set is shown in the table:

Table 1. Investigation and analysis of crop diseases and insect pests

Category number	Plant category	Type of illness	Numerical
1	Apple	Healthy	200
2	Apple	Cedar Rust	200
3	Banana	Healthy	200
4	Banana	Black leaf streak	200
5	Banana	speckle	200
6	Blueberry	Healthy	200
7	Cabbage	Healthy	200
8	Cassava	Brown leaf spot	45
9	Celery	Blight	200
10	Cherry	Powdery milder	200
11	Corn	leaf spot	200
12	Corn	Common rust	200
13	Corn	Healthy	200
14	Cucumber	Downy mildew	200
15	Cucumber	Healthy	58
16	Gourd	Downy mildew	166
17	Grape	leaf blight	200
18	Grape	Healthy	200
19	Grape	Esca	200
20	Onion	Healthy	58
21	Pepper	Healthy	200
22	Potato	Healthy	200
23	Potato	Early blight	156
24	Pumpin	Mosaic	200
25	Paspberry	Healthy	200
26	Soybean	Healthy	200
27	Soybean	Frogeye	200
28	Squash	Powdery mildew	154
29	Squash	Healthy	200
30	Strawberry	Leaf scorch	200
31	Strawberry	Healthy	200
32	Tomata	Healthy	200
33	Tomata	Septoria	200
34	Tomata	Late blight	200
35	Tomata	Bacterial spot	200
36	Tomata	Mosaic virus	200
37	Tomata	Eary blight	200
Total	-	-	6837

Based on the information in the above table, the preprocessing of visible light images mainly includes redefining the size of the images, de-averaging the images, and rewriting the data format. The sizes of the original visible light pictures are all around 800*500*3 pixels, and the sizes of the picture samples are different. For the research situation of, the visible light picture has too many pixels, contains a large amount of redundant information, and brings a relatively large burden to the calculation process [5]. Therefore, the dimension redefinition changes the size of all pictures to 224*224*3 pixels through downsampling, realizing normalization of all picture feature samples of crops. After redefining the sample characteristics, the sample set of crop diseases and insect pests is displayed and judged, as shown in the figure:

ISSN: 2307-0692, Volume 8, Issue 4, August, 2019

Figure 1. Sample collection of crop diseases and insect pests

The characteristic values of crop diseases and insect pests are redefined based on the sample set of crop diseases and insect pests, and further analyzed according to the results of the above figure. After multi-sampling treatment, the de-averaging operation is carried out. The specific method is to subtract the mean values from R, G and B. Then, due to the limitation of the selected machine learning framework, the redefined RGB image needs to be converted into 4-dimensional data [6]. Therefore, the second step of this part of work is to expand the dimensions of RGB images. In order that the expanded dimension does not affect the classification of data, the fourth dimension of each color image is set to 1. The generated 4-dimensional data is used as input for machine learning. Relying on the idea of feature extraction and SVM classification, the specific method is as follows:

Processing the data in advance, including redefining the size of the picture and enhancing the contrast of the picture;

Turn the processed picture into a gray image, and use the binarization method to turn it into a picture containing only white and black dots. After that, the color mode of the picture is changed from RGB to HSI；;

K-means clustering is carried out on the samples processed in the previous step, and segmentation is carried out on the samples [7].

Extracting features from the processed samples with a dimension of 13. Then, the 13-dimensional features are used to represent the original images for comparison.

13-dimensional features of each picture sample are taken as input to an improved support vector machine model to calculate classification accuracy rate

2.2. Intelligent detection algorithm for crop diseases and pests

The intelligent detection algorithm of crop diseases and insect pests is further optimized based on feature samples. For feature training of collected samples, in order to obtain a better and more mature detection model, a large number of feature values of collected samples need to be marked so as to correctly analyze the image sample set. However, due to the limited number of crop pest and disease picture samples currently available, if the network is trained directly with the obtained picture sample set, the training effect may not be ideal [8]. However, training methods based on transfer learning and machine learning will provide some help to solve this problem. The accuracy and high reliability of classification models are very important indicators in both traditional classification learning and modern classification learning represented by machine learning principles. Based on the above indicators, the extraction of color features of plant diseases and insect pests regions is mainly carried out in RGB space. In order to ensure the stability of the scale, the first-order and third-order color moments are calculated, and the normalized color histogram is calculated, i.e. the color features of plant diseases and insect pests. The specific algorithm is as follows:

$$\mu = \frac{1}{2N} \alpha_{\mathrm{i}=1}^{N-1} p \tag{1}$$

$$\sigma = \sqrt{\frac{1}{2N} \alpha_{j=1}^{N-1} (p+1)^2} \tag{2}$$

$$s_t = \sqrt{\frac{1}{2N} \alpha_{\mathrm{i},j=1}^{N} p * \frac{H_c(N_{\mathrm{ij}})}{w'h}} \tag{3}$$

In the above formula, p is used to describe the detection characteristic value of the I-th crop, where the gray scale of the sample channel is j, the occurrence probability of plant diseases and insect pests is n, and h is used to describe the histogram of the plant characteristic sample a. I and j are respectively used to describe the standard characteristic values and difference values of samples collected from plant samples [9]. W' is used to describe the regional moment order of plant diseases and insect pests, and H is used to describe the inertia product of crop detection samples. The sphericity, eccentricity and moment invariants of plant diseases and insect pests regions are calculated as grouping feature parameters to complete further grouping of morphological features. The specific algorithm is as follows:

$$\mathrm{k} = \frac{\mu r_i}{2 s_t} \tag{4}$$

$$E = \sqrt{\frac{k\sigma_{ij} - z r_i}{2(x+y+z) + \sum (A+B+C)^2 + \Delta k}} \tag{5}$$

In the above algorithm, ri is used to describe the inscribed circle and circumscribed circle of plant diseases and insect pests. A, b and c are respectively used to describe the rotation vectors of rigid bodies around (x, y, z) coordinate axes. In the inspection process, the texture features of plant diseases and insect pests are the key indicators for judging the types of plant diseases and insect pests, and the gray level co-occurrence matrix is the embodiment of the key texture features of the image. Therefore, the texture features of crop diseases and insect pests are calculated by using the angle second order moment, contrast and correlation, and the algorithm is as follows:

$$ASM = E - \frac{1}{2s_t} \tag{6}$$

Based on the above algorithm, weak classification processing is carried out on a plurality of collected samples to construct a strong classification set, which is recorded as follows ASM_m (m=1,2,⋯, M) If the actual collected sample data in the detection process is: $(a_1,b_1),\cdots,(a_m,b_m), b \in \{-1,1\}$. Under the above circumstances, the characteristic samples of crop data are normalized, and the weight of the processed numerical samples is $D(i) = \frac{1}{N} ASM_m$. Based on the above algorithm, the numerical value of the degree of influence of crop spectral information on plant diseases and insect pests is finally calculated, and the algorithm is as follows:

$$A_m(x)' = \frac{1}{2}\log ASM \sum_{i,j=1}^{m} E - \lim_{x\to\infty} k\sqrt{\mu^2 - s_t} \tag{7}$$

Based on the above algorithm, the vegetation index feature vector of spectral features is normalized, and the similarity between the sample points, normal samples and disease samples is calculated. If in the detection process, the average value of feature vectors representing healthy samples of crops represents the feature vectors of any sample. Then the difference of similarity of disease samples is calculated to obtain:

$$\Im = \sum^{\delta} \lim_{x\to\infty} ASM * \sqrt{A_m(x)\left|F^H - F^G\right|} \tag{8}$$

Based on the above algorithm, the impact degree of crop diseases and insect pests is calculated and analyzed, so as to better detect the growth of crops.

2.3. Realization of intelligent detection of crop diseases and insect pests

Both in traditional classification learning and modern classification learning represented by convolutional neural network are very important indicators. However, in traditional classification problems, the following two basic assumptions are usually used to ensure that the trained classification model has the above two properties:

Whether in the training set or the test set, the condition with the same probability distribution needs to be satisfied.

The number of training samples needs to be guaranteed, that is, there should be enough valuable training samples for training, so that a mature classification model can be trained.

In view of the above problems, combined with the previous algorithm to further carry out intelligent pest detection [10]. For machine learning, although the number of crop feature training samples collected is limited, it cannot be completely dependent on detection training, and needs to be corrected by means of unsupervised learning, parameter fine tuning and other methods, so as to achieve better adaptability. However, when there are a lot of data sets available for training, the details of supervised training are given below:

Randomly selecting samples in the original data set to obtain a training sample set;

The network parameters are randomly assigned, and the magnitude of the assignment depends on the situation. The learning rate η and the threshold of sample error "are given initial values.

Selecting one or more samples from the training sample set and inputting the samples into the network, and recording the real label classification thereof;

Use the network to calculate the sample and obtain the classification accuracy. Comparing the calculated classification results with the real labels to obtain the differences between their judgments;

Using gradient descent to spread the difference in the opposite direction. The propagation direction is the minimization direction.

After the back propagation is completed, the classification results of the input sample set are calculated again to obtain new errors. If the error value e meets the requirements, the training is stopped.

After the training is completed, fix the parameters and apply them to the actual situation. Machine learning has great advantages for image detection. When the posture of the object changes or the shape changes, the recognition ability of machine learning to the object is still not affected.

Compared with the classification method based on traditional neural network, machine learning has the following advantages:

Convolution neural network has advantages in recognition when some changes occur in image samples. For example, convolutional neural network has a good generalization effect for image rotation and scaling.

Weight sharing technology is another advantage of convolutional neural network. It can effectively reduce the computational complexity and optimize the parameter structure inside the convolutional neural network, thus realizing stronger adaptability.

The weak classifier is established according to the difference between the similarity between the sample and the disease sample and the similarity between the sample and the healthy sample, because the main reason for the overfitting problem is that the number of training samples is too small and the learning ability is too strong due to the too complex network structure. Therefore, a simpler structure can effectively alleviate overfitting. The specific fitting detection structure is shown in the following figure:

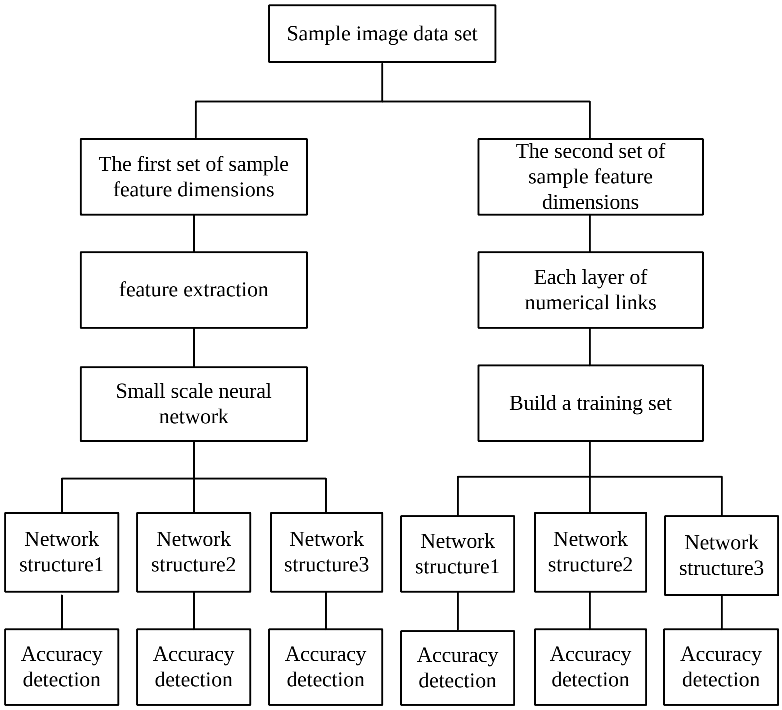

Figure 2. Monitoring method of crop diseases and insect pests

Based on the above method, the characteristics of crop diseases and insect pests are classified so as to more reasonably improve the accuracy and high reliability of the detection results, thereby better realizing the goal of timely and effective detection of crop diseases and insect pests.

3. Analysis of Experimental Results

In order to verify the rationality of the intelligent detection method of crop diseases and insect pests based on machine learning, a comparative detection experiment was carried out. Using the same data set of diseases and insect pests collected in Table 1, a convolution network of crop sample characteristics was built, and the characteristics of plant diseases and insect pests were trained. Here, the pest data set is divided into two types, the first is training sample set for machine learning training and the second is test set for analyzing the classification performance of convolution network. The proportion of different training samples to the total number of samples is set at 0: 10: 9, and then experiments are carried out for each different proportion. The specific control results of training samples are shown in the following table.

Table 2. Resolution value of crop diseases and pests

Proportion of training samples	Classification accuracy
0.1	0.377
0.2	0.566
0.3	0.612
0.4	0.542
0.5	0.781
0.6	0.645
0.7	0.578

From the experimental results, it can be seen that the machine learning modeled on Alex-Net for the applied image sample data set plays a certain role in the detection of crop diseases and insect pests, which is greatly improved compared with the detection method of feature extraction and SVM. This can reflect the advantages of detection based on machine learning. However, it can be seen from the table that training machine learning alone cannot achieve very high accuracy in pest detection. At present, the detection rate is kept below 80%, which is not of practical value. At the same time, it can be seen that the ratio of picture sets used for training has a relatively important relationship with the detection performance. As the proportion of training samples to all samples gradually increases, the detection accuracy of machine learning for crop diseases and insect pests gradually increases, and the improvement effect is obvious. After analysis, we can know that the machine learning layer modeled on Alex-Net is more and the structure is more complex, so it is easy to produce over-fitting problems. Over-fitting is a very common problem in the field of machine learning. The main reason is that the training samples are too few, which leads to incomplete knowledge acquired in the process of training and learning the training samples. Therefore, the classification results of the test samples are deficient. Therefore, solving the over-fitting problem will become the main research direction in the following part.

Table 3. Standard values of experimental parameters

Species of plant diseases and insect pests	Percentage of training samples	Classification parameter%	Standard value%
5	0.6	51.81	91.25
5	0.7	55.48	90.45
10	0.8	59.45	90.45
10	0.7	61.84	92.48
37	0.8	66.48	91.84
37	0.8	57.94	93.45

On the basis of the above experimental parameters, the support vector machine is used as the classification module instead of the original softmax classification layer, which can improve the classification accuracy to a small extent, alleviate the over-fitting problem to a certain extent, and further carry out comparative detection on the accuracy of the detection results under this environment. the specific experimental detection results are shown in the following figure:

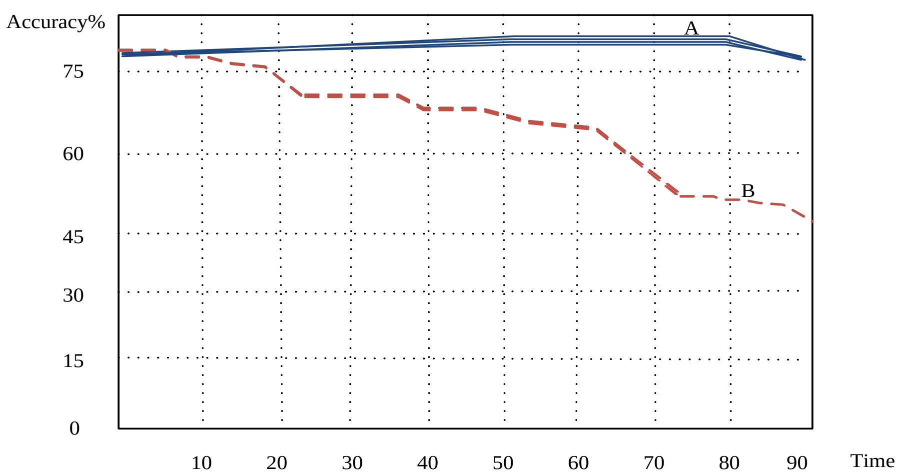

Figure 3. Comparison test results

In the above experiments, experimental curve A represents the detection results of crop diseases and insect pests based on machine learning, and B represents the detection results of traditional methods. After observing the above experimental results, it is not difficult to find that in the initial stage of the experiment, the detection results of the two methods are basically consistent. After 10 seconds, the experimental detection curve B shows an obvious downward trend, while the curve A remains stable all the time. This proves that compared with the traditional methods, the machine learning-based detection method for crop diseases and insect pests proposed in this paper has higher accuracy and stability, and fully meets the research requirements.

4. Conclusion

In this paper, a visible light image based on machine learning is proposed to detect crop diseases and insect pests. Its main function is to classify crops through machine learning, so as to realize the detection of diseases and insect pests. This paper introduces the detection results of data sets based on machine learning only. The detection accuracy is not very high and there is a serious over-fitting problem. Then, in order to alleviate the overfitting phenomenon, the idea of transfer learning was introduced. Through designing machine learning with different structures and training based on the idea of transfer learning, the over-fitting problem is alleviated to a great extent and the detection accuracy of the detection method is improved.

5. Acknowledgment

The research is supported by Hainan Provincial Natural Science Foundation of China (No. 619QN243).

References

[1] Estimation of corn leaf area index based on GF-1 WFV image and machine learning algorithm. Journal of Image and Graphics. 2018, 23(5), 719-729.

[2] Liu Zhong, Wan Wei, Huang Jinyu. Research progress on inversion of key parameters of crop growth based on uav remote sensing. Journal of Agricultural Engineering. 2018, 34(24), 68-79.

[3] Chen Jiguang, Song Xiandong, Wang Chunrong, et al. Development and application of heilongjiang crop pest online monitoring and management system. China Plant Protection. 2017, 37(8), 26-32.

[4] Zhang Xinliang, Xia Yafei, Xia Nan, et al. Design of crop pest monitoring system based on embedded image terminal. Software Guide, 2018, 17, 189(7), 128-131.

[5] Huang Yanqing. Research on development of intelligent decision support system for crop diseases and pests in northern liaoning. Agricultural Science and Technology and Equipment. 2017, (2), 24-25.

[6] Zheng Qingwei. National agricultural technology center organized field application investigation and exchange of intelligent pest collection equipment. Pesticide Market Information. 2018, 622(13), 16.

[7] Liang Xiaotong, Xu Jian. Application analysis of MATLAB image processing technology in identification of agricultural diseases and pests. South China Agriculture. 2017, 11(21), 117-122.

[8] Li Jian, Liu Na, Cheng Ling, et al. Application of information technology in monitoring and early warning of crop diseases and pests. Jiangsu Agricultural Sciences. 2018, 46(22), 9-14.

[9] Xena, Yang Siqing, Xie Jun. Design and implementation of early warning information system for rice diseases and pests based on WebGIS. Value Engineering. 2017, 36(35), 183-185.

[10] Wu Zongzhuo. Application of intelligent identification and diagnosis technology of rice diseases and insect pests in disease and insect pest monitoring. Modern Scientific Instruments. 2018, (3), 87-90.