


Research Article

A Recognition Method of Crop Diseases and Insect Pests Based on Transfer Learning and Convolution Neural Network

Dongmei Liu,^{1,2} Hangxu Yang ,^{1,2} Yongjian Gong,^{1,2} and Qing Chen³

¹College of Mechanical and Electronic Engineering, Jinhua Polytechnic, Jinhua 321017, Zhejiang, China

²Key Laboratory of Crop Harvesting Equipment Technology of Zhejiang Province, Jinhua 321017, Zhejiang, China

³College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing 210037, China

Correspondence should be addressed to Hangxu Yang; yanghx_83@126.com

Received 26 May 2022; Accepted 22 June 2022; Published 5 July 2022

Academic Editor: Zaoli Yang

Copyright © 2022 Dongmei Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The research on crop pest identification technology based on image analysis has important practical significance for effectively controlling the occurrence of crop diseases and insect pests (CDIP), improving crop yield and reducing the pollution of pesticides to the environment. Aiming at the problems existing in crop pest recognition technology and aiming at improving the accuracy and efficiency of crop pest recognition, this study proposes a crop pest image recognition method based on transfer learning and convolution neural network. Firstly, the crop pest image is geometrically operated to expand the crop pest image data set, and then the expanded data set is divided into training set and test set. Then, Alexnet, VGG-16, and ResNet-50 models are pretrained on ImageNet large image data set. Based on the theory of transfer learning, the learned model parameters are transferred to the small sample image data set of crop diseases and insect pests in this paper, so as to train the crop diseases and insect pests recognition model. This method is tested on Image Database for Agricultural Diseases and Pests (IDADP). The experimental results show that the convolution neural network plant leaf image recognition method based on transfer learning has better effect. It can quickly and accurately diagnose crop diseases, reduce the use of pesticides and fertilizers, and improve the yield and quality of crops.

1. Introduction

Diseases and insect pests directly affect the yield of crops in agricultural production. It is of great significance to timely and effectively control the crop diseases. The research purpose of this paper is to apply the artificial intelligence technology in the field of computer, aiming at the problems existing in the identification technology of crop diseases and insect pests, and aiming at improving the accuracy and efficiency of crop diseases and insect pests identification, detect the plant diseases and insect pests image, identify the diseases and insect pests of crops, so as to carry out targeted drug use for the diseases and insect pests crops, and reduce drug waste and pollution. In the process of crop planting, it is often affected by various adverse environmental factors, which destroys its metabolic process, leads to the occurrence of crop diseases, makes crops unable to develop normally, causes losses to agricultural production, and indirectly

affects crop yield and product quality [1, 2]. Due to the long cycle of biological control and ecological control, the current pest control method is still through chemical control. The causes of different diseases and pests of crops are different, so the pesticides used in prevention and control are also different [3]. The long-term extensive use of chemical fertilizers and pesticides has caused serious pollution to the environment. Therefore, the correct identification of CDIP plays an important guiding role in the prevention and control of CDIP [4]. In the actual production process, agricultural workers encounter the following problems when identifying crop diseases: the initial symptoms of CDIP are often vague, and ordinary agricultural workers lack comprehensive knowledge of CDIP diagnosis. They usually spray a large amount of pesticides on crops when the incidence of CDIP is serious [5]. Although crop protection experts have a lot of knowledge on diagnosis of plant diseases and insect pests, due to their busy work and large service area, they do not

have time to often diagnose CDIP for the majority of farmers and provide consulting services for the prevention and control of CDIP for the majority of farmers [6]. The symptoms of CDIP are complex and fuzzy. The traditional diagnosis methods of CDIP include agricultural diseases and insect pests Atlas method, agricultural diseases and insect pests method, and classification key table method. These methods are difficult for ordinary agricultural workers to master certain professional knowledge of agricultural diseases and insect pests [7]. Moreover, the description of symptoms adopts language and words, which is fuzzy, so the detection method is time-consuming and laborious; it seriously affects the accuracy of pest prediction.

In recent years, with the development of information technology and the improvement of people's living standards, agricultural workers can easily obtain and process the pictures of crop diseases from outdoors. The application of new agricultural production technologies such as "precision agriculture" and "digital agriculture" has made the traditional agricultural production develop towards automation and intelligence [8]. Using high-tech means such as image processing technology and pattern recognition technology to judge the damage status of CDIP, and then take corresponding disease control measures in time can reduce human and financial resources, which is conducive to the intelligent development of crop cultivation management [9]. Image processing and pattern recognition technology are widely used in agricultural robots, precision agriculture, and other fields, which can be used as a good reference for the research on the diagnosis and recognition of CDIP [10]. Through in-depth learning, the pest image is segmented, effective features are extracted and classified, so as to quickly and accurately identify the types of pests and diseases, so as to realize the automatic and intelligent identification of CDIP [11]. If the image processing technology and pattern recognition technology can be used to automatically and accurately identify CDIP, and then combined with the biological characteristics such as texture, shape, and color of the disease spot image when the crop is infected, the disease types of CDIP can be quickly identified, so as to reduce the loss of agricultural producers caused by agricultural diseases and reduce the pollution of pesticides to agricultural products and the environment. It is of great significance to improve the economic benefits [12].

Traditional feature-based CDIP recognition methods need complex preprocessing and feature extraction operations, and there is no unified conclusion on the quality of the extracted features. The convolution neural network can directly take the crop image as the input of the network, and can automatically extract the image features of CDIP without preprocessing [13]. Moreover, the features extracted by the convolution neural network have stronger discrimination and generalization ability than the manually designed features. However, the training of deep convolution neural network model usually requires a large number of labeled samples. Without sufficient training data, the model is likely to fall into overfitting. Only when the network structure is complex and the number of training samples is large enough, the convolutional neural network can show

excellent performance, but it is very difficult to obtain large-scale labeled data at present. The proposal of transfer learning solves the above problems caused by the lack of training samples, and has been widely used in the field of image recognition. Considering that the CDIP database belongs to a small sample database, based on the transfer learning theory, this paper applies the pretrained AlexNet, VGG-16, and ResNet-50 models to the classification of CDIP images [14, 15].

2. Related Work

With the perfection of computer vision technology, using image processing technology to realize image segmentation, feature extraction and classification and recognition of CDIP has become the main research direction. In recent years, scholars have conducted research on CDIP image segmentation technology, feature extraction technology, and disease diagnosis and recognition technology, and have made many achievements, but there are still some problems that need to be further studied and improved, mainly in the following aspects: the external symptoms of crop leaves are relatively complex, using the color, texture, and single characteristic parameters such as shape cannot better describe the characteristics of CDIP, which will affect the correct identification of CDIP. Traditional pattern recognition methods are used in most crop leaf disease recognition. For complex crop disease images, it is difficult to obtain ideal training patterns and classification results. Aiming at the problems existing in the research of CDIP identification technology, aiming at improving the accuracy and efficiency of CDIP identification, this study developed a good interactive crop leaf disease model [16, 17]. Taking CDIP as the research object, this paper studies the key technologies such as image segmentation, feature extraction, disease diagnosis, and recognition of crop leaf diseases and pests, so as to diagnose CDIP quickly and accurately, reduce the use of pesticides and fertilizers, and improve crop yield and quality. The application fields of machine learning include pattern recognition, data mining, statistics, computer vision, speech recognition, and natural language processing [18, 19]. The concept of deep learning comes from the research of artificial neural network, but its development solves many problems that traditional artificial intelligence cannot solve, and speeds up the application of artificial intelligence in real society. The deep learning method is divided into supervised learning method and unsupervised learning method, which mainly includes three steps: forward propagation, loss calculation, and back propagation [20]. Deep learning involves knowledge in many fields, such as linear algebra, probability theory, and biology. In recent years, scholars have applied various deep learning structures to crop phenotype analysis and pest identification, including DCNN, RCNN, GoogleNet, ResNet, SegNet, AlexNet, VGGNet, ResNeXt, SqueezeNet, GAN, LeNet18, and ZFNet. At present, deep learning structure performs well in a wide range of plant phenotypic classification tasks, such as plant recognition based on leaf vein pattern, leaf count, tassel count in maize, ear segmentation, root and stem location

and feature detection, flowering detection, and plant recognition [21].

With the development of computer vision technology, the application of computer image processing technology and pattern recognition technology in agricultural pest identification is gradually carried out. Combined with the external characteristics of the color, texture, and shape of crop diseases and pests, the image processing technology and pattern recognition technology are used to recognize the crop disease image [22]. While obtaining the leaf image of diseases and pests, the knowledge of the type, degree, and control method of diseases and pests can be displayed to the user in real time. Therefore, the diagnosis of crop leaf diseases based on computer vision technology is effective. The research on crop disease recognition can be divided into several aspects: disease spot image segmentation; Disease image feature extraction; Disease image classification and recognition. Machine learning is a method based on a large amount of data, using the existing massive data to analyze and infer a certain model and use it to predict the future [23]. It is divided into supervised learning, unsupervised learning, and reinforcement learning. The classical algorithms of machine learning include decision tree algorithm, artificial neural network algorithm, support vector machine algorithm, and naive Bayesian algorithm. Deep learning is the extension of neural network algorithm in machine learning. Depth refers to the number of layers of neural network [24, 25]. It only solves the disadvantage of single-layer learning machine and medium-level learning machine. With the development of machine learning, it will automatically learn the high-level features implied in the data, and gradually improve with the improvement of the model and the expansion of training data, so as to develop deep learning. Common neural networks for deep learning include recurrent neural network (RNN), Convolutional Neural Network (CNN), Restricted Boltzmann Machine (RBM), and Deep Belief Network (DBN) [26, 27].

There are also a lot of applications in plant phenotype detection under pest stress. Mohanty et al. [28] used 54306 images to train DCNN (AlexNet and GoogleNet Architecture) to identify 26 diseases from 14 different crop species. The training model using color images in GoogleNet achieved 99.35% accuracy. Ferentinos [29] used VGG CNN and achieved 99.5% success rate in detecting plant phenotypic stress. Amanda et al. [30] trained the INVONCE-v3 model for five cassava diseases. The results showed that the recognition accuracy of three cassava diseases (cassava brown strip, cassava mosaic disease, and brown spot) and two mites was good, and the average accuracy of the test data was 93%. Lu et al. [31] deployed a new architecture called deep multipoint learning (DMIL) on the top of the wheat disease database (WDD2017). The database contains 9230 images of 6 common wheat diseases and 7 different categories of healthy wheat. Among them, VGG-FCN is better than VGG-CNN in the recognition accuracy of disease categories. Ubbens et al. [32] used artificial neural network and support vector machine as pattern recognition methods to classify insects, and designed a new automatic insect specimen image recognition system, with the stability and accuracy of 93%.

3. Identification of CDIP Based on Transfer Learning and Convolution Neural Network

3.1. Three Convolutional Neural Network Structures

3.1.1. AlexNet Network Structure. The AlexNet model consists of five convolution layers and three pooling layers, including three fully connected layers. The structure of AlexNet is similar to that of LeNet, but more convolution layers and more parameter space are used to fit the large-scale data set ImageNet. It is the boundary between the superficial nerve collaterals and the deep nerve collaterals. AlexNet model adds ReLU activation function after each convolution, which solves the gradient disappearance problem of sigmoid and makes the convergence faster. The random discard technique (dropout) is used to selectively ignore a single neuron in the training to avoid overfitting of the model (data enhancement is also used to prevent overfitting). A normalized LRN (Local Response Normalization) layer is added to improve the accuracy. Overlapping Max pooling, that is, the pooling range z is related to the step size s . $z > s$ avoids the average effect of average pooling.

AlexNet successfully uses ReLU as the activation function of CNN, which verifies that its effect exceeds sigmoid in deeper networks, and successfully solves the gradient dispersion problem of sigmoid in deeper networks. During training, dropout is used to randomly ignore some neurons to avoid overfitting of the model. It is generally used in the full connection layer. Dropout is not used in prediction, that is, dropout is 1 overlapping maximum pooling (step size less than convolution kernel) is used in CNN. Previously, CNN generally used average pooling, and the use of maximum pooling can avoid the fuzzy effect of average pooling. At the same time, the overlapping effect can improve the richness of features. The LRN (Local Response Normalization) layer is proposed to create a competition mechanism for the activities of local neurons, so that the value with large response becomes relatively larger, inhibit other neurons with small feedback, and enhance the generalization ability of the model.

3.1.2. VGG-16 Network Structure. VGG-16 has a total of 16 layers, including 13 convolution layers and 3 fully connected layers. Layers are separated by Max pooling, and the activation units of all hidden layers adopt ReLU function. Using the pool layer as the boundary, VGG16 has 6 block structures, and the number of channels in each block structure is the same. Both convolution layer and full connection layer have weight coefficients, and the pooling layer does not involve weight. For VGG16 convolution neural network, convolution layer and pooling layer are responsible for feature extraction, and the last three full connection layers are responsible for completing the classification task. VGG uses the convolution layer of multiple smaller convolution cores ($3 * 3$) to replace the larger convolution layer of one convolution core. On the one hand, it can reduce the parameters; on the other hand, it is equivalent to more nonlinear mapping, which can increase the fitting/

expression ability of the network. All convolution layers are convolution cores of 3×3 , and the pool core of pool layer is 2×2 . VGG16 is a large network with 138 million parameters in total. Therefore, its main disadvantage is that the number of features that need to be trained is very large.

3.1.3. Resnet-50 Network Structure. In convolutional neural network, each layer of neurons is only connected with the left and right adjacent neurons, that is, the input data is received from the previous layer and transmitted to the next layer. In this way, the connection between adjacent layers not only restricts the diversity of the network, but also easily leads to the deepening of the number of network layers and the difficulty of model training. To solve this problem, ResNet network solves the problem of the decline of deep network training efficiency by means of cross-layer connection and fitting error term. Residual connection module is the core architecture of ResNet network, and identity mapping is its main idea. In the process of forward propagation, the network without residual module obtains the final output result of the network after convolution, pooling, and other operations. ResNet uses a global mean pooling layer. Using the residual module, the residual network of 152 layers can be trained. ResNet-50 mainly uses 3×3 convolution. ResNet-50 has two basic blocks, one is identity block. The input and output dimensions are the same, so multiple blocks can be connected in series, another basic block is Conv Block. The input and output dimensions are different, so they cannot be connected in series. Its function is to change the dimension of the eigenvector.

3.2. Transfer Learning Network Model. Transfer learning is a machine learning method that uses existing knowledge to solve problems in different but related fields. Its goal is to complete the transfer of knowledge between related fields. For convolutional neural networks, transfer learning is to successfully apply the “knowledge” trained on specific data sets to new fields.

There are usually two methods to apply transfer learning to convolutional neural networks. One is to use the pre-training model with learning weight to obtain the features to be used in the new problem, that is, the pretraining network model is used as the feature extractor. At this time, the output of the network extracts the features of interest in the part in front of the last full connection layer. The other is to use the new data set to train the network to fine-tune the network weight. In this case, the number of output layer nodes must be modified to match the number of categories in the new problem. In addition, in both cases, the input data must match the input size of the pretraining network. The specific transfer learning method depends on the size difference between the target data set and the original data set and the similarity between them. If the target data set is very small and they are similar, in order to prevent overfitting, the pretraining model is usually used as a feature extractor, on the contrary, fine-tuning is adopted.

In this paper, the above three deep convolution neural network models are used for pretraining on the large image

data set of ImageNet. Based on the theory of transfer learning, the learned model parameters are transferred to the small sample image of CDIP in this paper. On the basis of not changing other parameters, the number of neurons in the last full connection layer is replaced by the number of crop disease species. Because the data set used in this paper is very small and the possibility of overfitting is relatively large, the trained network model is used as a feature extractor, that is, the depth features of CDIP images are extracted with the trained model, and finally the extracted features are put into the Softmax classifier for classification training. The model architecture is shown in Figure 1.

In the transfer learning model in Figure 1 above, autoregression and automatic encoder are the most commonly used self-supervised learning methods. Autoregression aims to predict the occurrence probability of the next word by using the previous word sequence (language model). The purpose of automatic encoder is to reconstruct the original data of damaged input sentences, such as masking a word in the sentence, or disturbing the word order. These self-supervised learning methods are used to learn the context-sensitive representation of words. When doing specific tasks, the purpose of fine-tuning is to use its labeled samples to adjust the parameters of the pretraining network. Take Bert (a popular pretraining model) as an example to judge whether the two sentences have the same semantics. The input is two sentences, and the corresponding coding representation of each sentence is obtained through Bert. The first hidden node of the pretraining model can be simply used to predict the probability that the two sentences are synonymous sentences. At the same time, an additional linear layer and Softmax need to be added to calculate the distribution of classification labels. The predicted loss can be transmitted back to Bert, and then the network can be fine-tuned, of course, can also design a new network for specific tasks and take the results of pretraining as its input.

3.3. Selection of Classifier. The classification result is the convolution feature obtained by the classifier through training and learning pool. The classifier selected in this paper is Softmax classifier. Softmax regression algorithm is to analyze the relationship between the probability of the dependent variable taking a certain value and the independent variable. In essence, it can solve the problem of multiple classifications. The output results represent the probability that the sample belongs to each attribute, and then select the corresponding category corresponding to the optimal probability from these results as the category of the sample prediction results. For the n classification problem, given a sample x , the conditional probability formula of Softmax regression calculation is,

$$P(y = c | X) = \text{softmax}(W_T^n X) = \frac{\exp(W_c^T X)}{\sum_{c=1}^n \exp(W_c^T X)}, \quad (1)$$

where, y represents the category label, and $y \in \{0, 1, 2, \dots, n\}$, W_c represents the weight of category c .

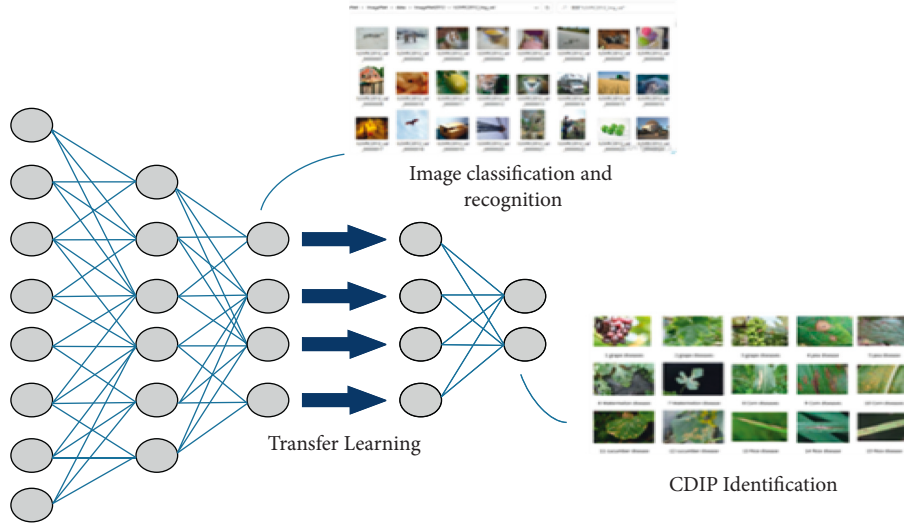


FIGURE 1: Architecture of transfer learning.

Assuming that the loss function is $J(\theta)$, $h(x)$ represents the function to be fitted, where θ is the parameter and m represents the number of all samples, the loss function expression is,

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^m (y^i - h_{\theta}(x^i))^2, \quad (2)$$

where, $h(\theta) = \sum_{j=0}^n \theta_j x_j$, n represents the number of features.

4. Experimental Test and Analysis

4.1. Data Set and Data Expansion. This paper constructs the construction of convolutional neural network model, the migration training and testing of pretraining model. The experiment uses IDADP to evaluate the performance of this method. IDADP provides CDIP image sample resources required for scientific research. IDADP is built by Hefei Institute of intelligent machinery, Chinese Academy of Sciences, together with Subtropical Agricultural Ecology Research Institute and Remote Sensing and Digital Earth Research Institute. It integrates CDIP image sample resources. Each kind of CDIP has hundreds to thousands of high-quality images; it can provide sample data for machine learning modeling for CDIP recognition research. Some images of CDIP are shown in Figures 2 and 3.

In order to reduce the overfitting phenomenon, each image is randomly flipped horizontally and vertically to expand the data set. Before using convolutional neural network to process the image, the sample is first de-averaged, and the formula is

$$x' = x - \mu, \quad (3)$$

where, x' is the sample after removing the mean value, x is the original sample, and μ is the sample mean value.

4.2. Experimental Setup and Evaluation Index. For the AlexNet model, the image size of CDIP is uniformly modified to $224 * 224$ pixels to train the IDADP data set without changing other parameters. Among them, the batch size of the network is set to 32, the learning rate is 0.001, the number of iterations is 5000, and the dropout is 0.5. The random gradient descent method is used for training.

For VGG-16 model, the image size of CDIP is uniformly modified to $227 * 227$ pixels to train the IDADP data set without changing other parameters. Among them, the batch size of the network is set to 64, the learning rate is 0.001, the number of iterations is 3000, and the dropout is 0.5. The random gradient descent method is used for training.

For ResNet-50 model, the image size of CDIP is uniformly modified to $224 * 224$ pixels, using the pretrained ResNet-50 model as the feature extractor to extract the feature vector of the IDADP data set, and then input the feature vector into the Softmax classifier to obtain the classification result. Among them, the batch size of the network is set to 128, the learning rate is 0.01, the number of iterations is 4000, and the dropout is 0.5. The random gradient descent method is used for training.

The accuracy of the test set is used to evaluate the quality of the experimental method. The calculation formula is as follows:

$$Acc = \frac{m}{n}, \quad (4)$$

where, m represents the number of correct test images, and n represents the total number of test set images.

4.3. Analysis of Experimental Results. As shown in Figures 4, 5, and 6, the curves of the accuracy and loss value of AlexNet, VGG-16, and ResNet-50 on the training set and test set with the number of iterations are shown, respectively.

As can be seen from Figure 4, there is a certain gap between the loss value of the training set and the test set of

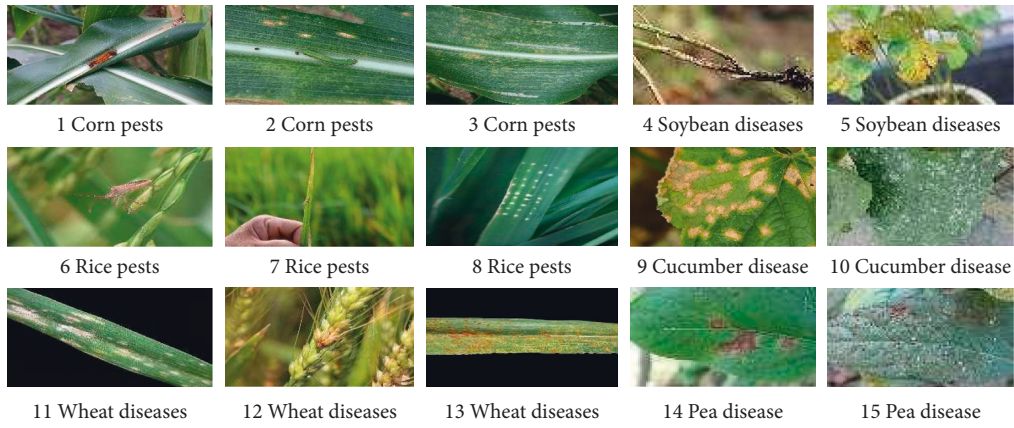


FIGURE 2: Sample pictures of CDIP.



FIGURE 3: Sample pictures of CDIP.

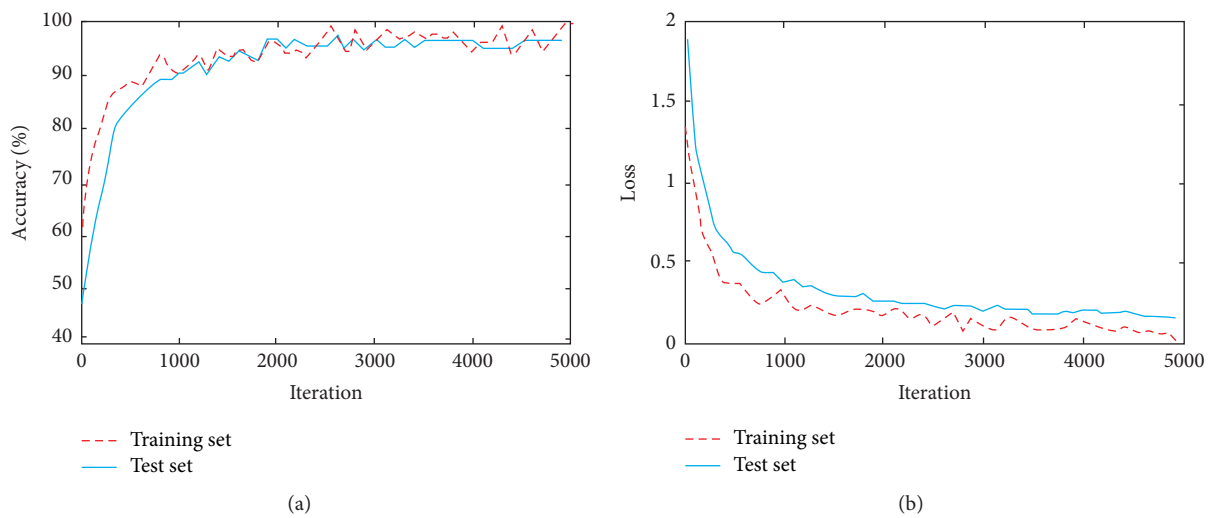


FIGURE 4: Accuracy and loss function curves of AlexNet (a) Accuracy value (b) Loss value.

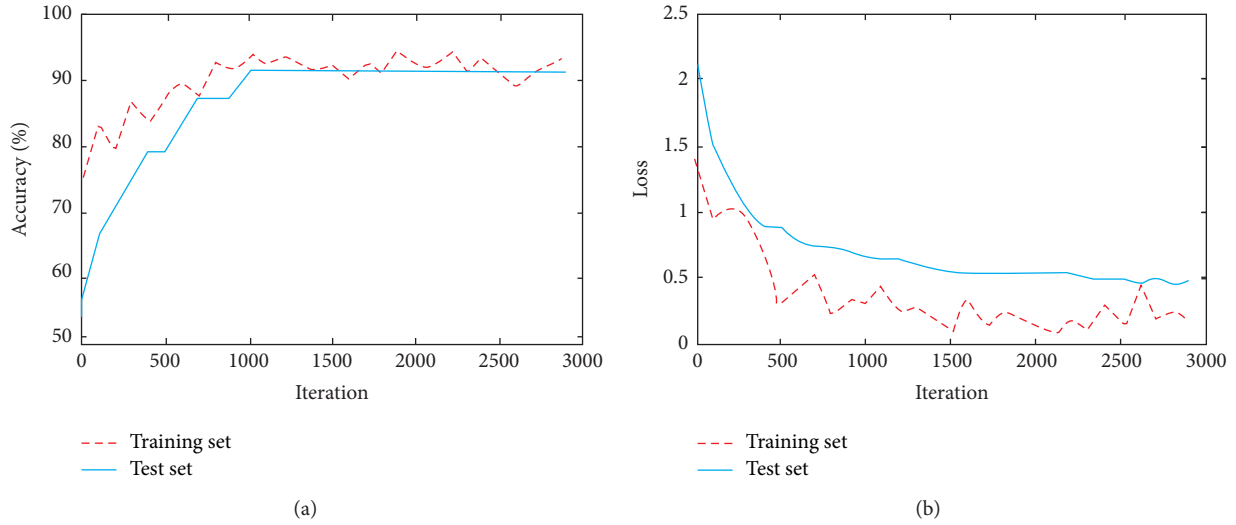


FIGURE 5: Accuracy and loss function curves of VGG-16 (a) Accuracy value (b) Loss value.

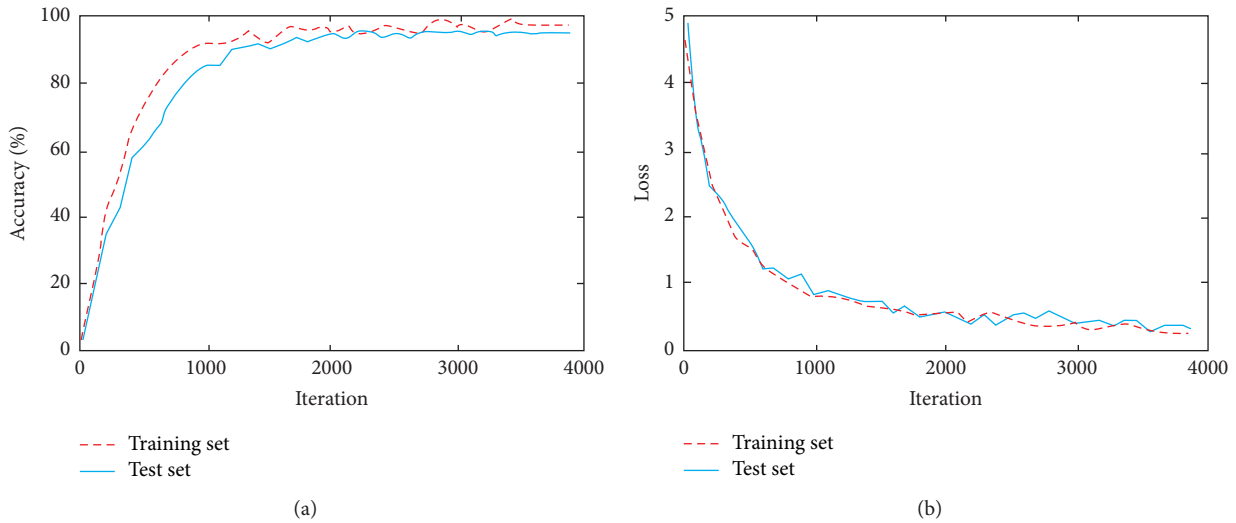


FIGURE 6: Accuracy and loss function curves of ResNet-50 model (a) Accuracy value (b) Loss value.

the AlexNet. When iterating for 2000 times, the loss value of the test set tends to be about 0.25 and hardly changes. As can be seen from Figure 5, when the VGG-16 is iterated for 1000 times, the accuracy curve of the test set will not change. At the same time, there is a certain gap between the loss value of the training set and the test set, and the loss value of the test set is about 0.53, resulting in overfitting. As can be seen from Figure 6, the change trend of loss curve of ResNet-50 on the test set and training set is basically the same, and the accuracy and loss value are basically stable when iterating 4000 times, and the convergence ability of the model is very good.

Based on the above, it can be seen from the accuracy and loss curve of the three models that ResNet-50 model performs better than the other two models in the small sample IDADP in this paper. In order to observe the influence of different gradient optimization algorithms on the accuracy of the test set, three gradient descent algorithms of SGD, moment, and Adam are used to train the three models,

respectively. The specific parameter configuration of the three gradient descent algorithms is shown in Table 1.

In Table 1 above, “—” indicates that this parameter is not used. The three models are trained with different gradient descent algorithms. The loss curve of the training set and the accuracy curve of the test set are shown in Figures 7, 8, and 9.

It can be seen from Figures 7, 8, and 9 that for the three models, different gradient descent algorithms are used for training, and the change trend of the loss curve of the three training sets is consistent with that of the accuracy curve of the test set. From the loss curve of the training set of the three models, it can be seen that in the early stage of decline, momentum, and Adam training methods have been well accelerated, while SGD training method is relatively slow. In the middle and late period of decline, the momentum term of momentum method increases the update amplitude and jumps out of the local minimum value, which can converge quickly. However, Adam algorithm oscillates repeatedly

TABLE 1: Specific configuration of different gradient descent algorithms.

Method	Momentum coefficient	First moment	Second moment
SGD	—	—	—
Momentum	0.913	—	—
Adam	—	0.946	0.982

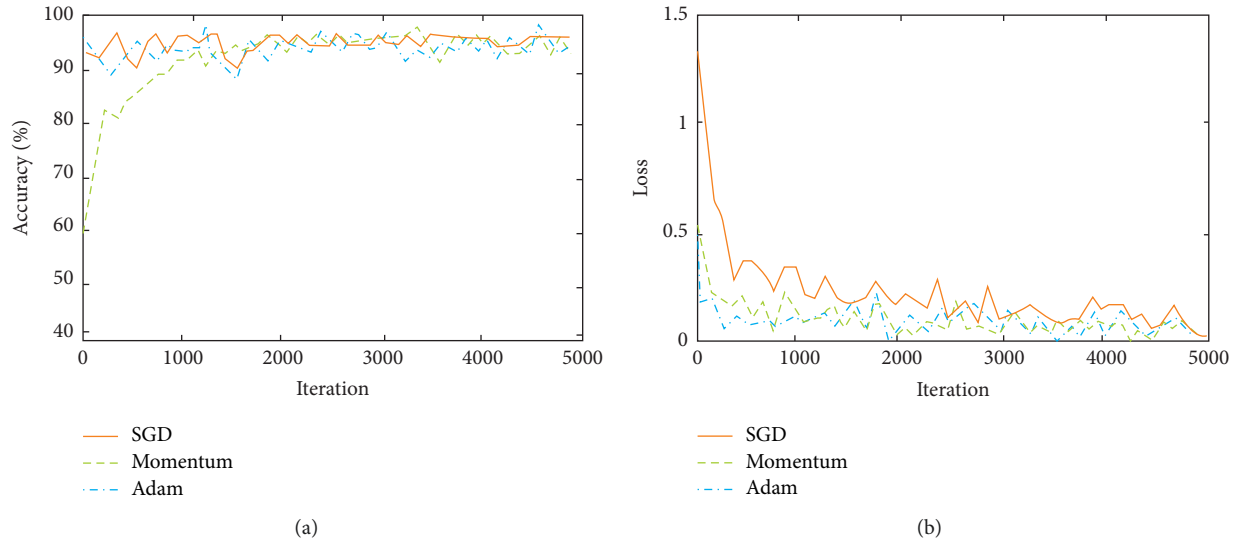


FIGURE 7: Accuracy and loss function curves of AlexNet model (a) Accuracy value (b) Loss value.

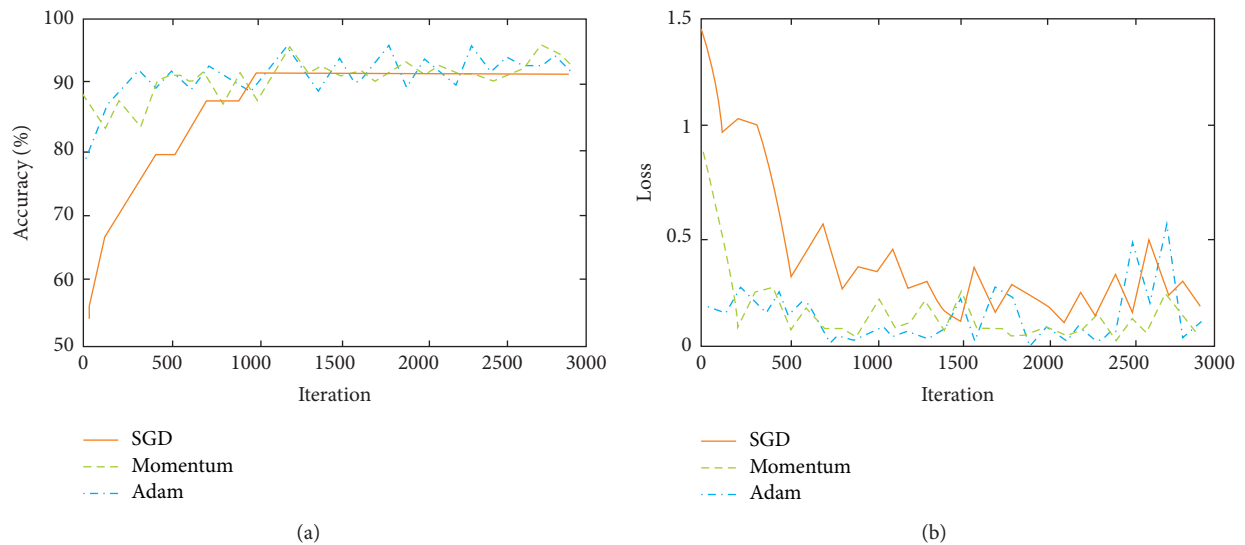


FIGURE 8: Accuracy and loss function curves of VGG-16 model (a) Accuracy value (b) Loss value.

near the local minimum value, which makes the function unable to converge. From the test set accuracy curves of the three models, it can be seen that the accuracy of Adam algorithm in the test set is higher than that of momentum algorithm, and the accuracy of SGD algorithm in the test set is the lowest, but there is little difference between the three. It should be noted that the momentum is helpful to accelerate convergence, but its role is almost negligible due to its small

learning rate and simple data set. This is why SGD is widely used in small sample data sets.

It can be seen that this is because the feature-based CDIP identification method mainly relies on manual extraction of crop features, and there is no unified conclusion about the quality of the features. Moreover, this method only extracts the crop shape features and texture features according to the previous experience to describe CDIP. This method uses the

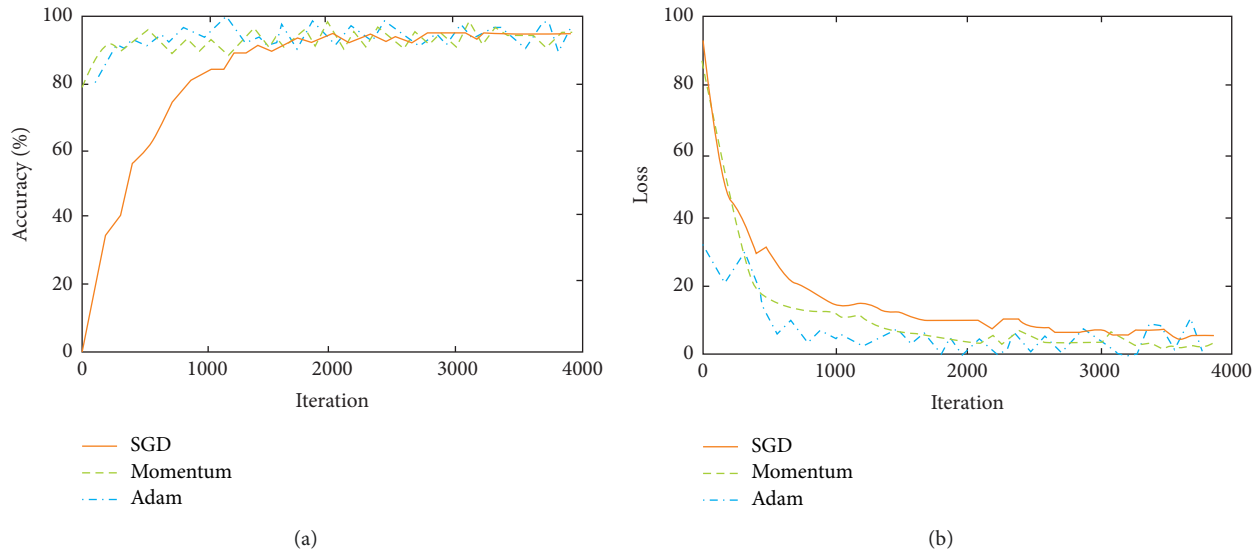


FIGURE 9: Accuracy and loss function curves of ResNet-50 model (a) Accuracy value (b) Loss value.

powerful feature extraction ability of migration learning and CNN to extract the image features of CDIP, and the extracted features are more comprehensive. Therefore, this method performs better in small sample CDIP image data set, which further shows that the convolution neural network method based on migration learning can be better applied to CDIP image data set with small sample size.

5. Conclusions and Future Work

In this paper, a method of CDIP recognition based on convolution image is proposed. This paper first introduces the three transfer learning models AlexNet, VGG-16, and ResNet-50 to be used in this paper, and then gives the corresponding model architecture. Considering that the image data set of CDIP in this paper belongs to small sample data set, this paper applies the above three deep convolution neural network models to pretrain on ImageNet large image data set, based on the theory of transfer learning. The learned model parameters are transferred to the small sample image data set of CDIP in this paper. Finally, experiments are carried out in the IDADP database. The accuracy of the three models in the training set and test set and the change of their loss value with the number of iterations are analyzed on the IDADP database, which verify the effectiveness of the method proposed in this paper.

The follow-up work will deeply study the structure of convolutional neural network, and study the influence of convolution kernel size, pool size, learning rate, iteration times, pool mode, and other network parameters on the convolutional neural network. At the same time, increase the scale and types of training samples to increase the accuracy of CDIP identification.

Data Availability

The authors confirm that the data supporting the findings of this study are available within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 51906111).

References

- [1] S. Aich and I. Stavness, "Leaf counting with deep convolutional and deconvolutional networks," in *Proceedings of the 16th IEEE International Conference on Computer Vision (ICCV)*, pp. 2080–2089, IEEE, Venice, Italy, October 2017.
- [2] G. L. Grinblat, L. C. Uzal, M. G. Larese, and P. M. Granitto, "Deep learning for plant identification using vein morphological patterns," *Computers and Electronics in Agriculture*, vol. 127, pp. 418–424, Sep 2016.
- [3] A. Ramcharan, K. Baranowski, P. McCloskey, B. Ahmed, J. Legg, and D. P. Hughes, "Deep learning for image-based cassava disease detection," *Frontiers of Plant Science*, vol. 8, p. 1852, 2017.
- [4] R. Xu, C. Y. Li, A. H. Paterson, Y. Jiang, S. P. Sun, and J. S. Robertson, "Aerial images and convolutional neural network for cotton bloom detection," *Frontiers of Plant Science*, vol. 8, Article ID 02235, 2018.
- [5] M. P. Atkinson, A. J. Townsend, M. H. Wilson et al., "Deep machine learning provides state-of-the-art performance in image-based plant phenotyping," *GigaScience*, vol. 6, no. 10, pp. 1–10, 2017.
- [6] H. Lu, Z. G. Cao, Y. Xiao, B. H. Zhuang, and C. H. Shen, "TasselNet: counting maize tassels in the wild via local counts regression network," *Plant Methods*, vol. 13, no. 1, p. 79, 2017.
- [7] M. Sulc and J. Matas, "Fine-grained recognition of plants from images," *Plant Methods*, vol. 13, p. 115, 2017.
- [8] X. Duan, L. Liu, H. Tu et al., "Panicle-SEG: a robust image segmentation method for rice panicles in the field based on deep learning and superpixel optimization," *Plant Methods*, vol. 13, no. 1, p. 104, 2017.

- [9] A. Fuentes, S. Yoon, S. C. Kim, and D. S. Park, "A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition," *Sensors*, vol. 17, no. 9, p. 2022, 2017.
- [10] A. P. Naufal, C. Kanjanaphachaoat, A. Wijaya, N. A. Setiawan, and R. E. Masithoh, "Top, "Insects identification with convolutional neural network technique in the sweet corn field," in *Proceedings of the 2nd International Conference on Sustainable Agriculture for Rural Development (ICSARD)*, vol. 653, p. 2021, Electr Network, Telangana, India, October 2020.
- [11] L. G. Tian and C. Liu, "Research on plant diseases and insect pests identification based on CNN," in *Proceedings of the 6th International Conference on Agricultural and Biological Sciences (ABS)*, vol. 594, p. 2020, Electr Network, Tokyo, Japan, December 2002.
- [12] P. Bhatt, S. Sarangi, A. Shivhare, D. Singh, and S. Pappula, "Identification of diseases in corn leaves using convolutional neural networks and boosting," in *Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods (ICPRAM)*, pp. 894–899, Prague, Czech, February 2019.
- [13] D. Hu, H. Shi, and W. Jiang, "Infrared infrared and visible image fusion based on empirical curvelet transform and phase congruency," *Ukrainian Journal of Physical Optics*, vol. 22, no. 3, pp. 128–137, 2021.
- [14] J. Chen, W. Chen, A. Zeb, D. Zhang, and Y. A. Nanekaran, "Crop pest recognition using attention-embedded lightweight network under field conditions," *Applied Entomology and Zoology*, vol. 56, no. 4, pp. 427–442, Nov 2021.
- [15] D. Defa, S. Hailiang, and W. Weijin, "Infrared and visible image fusion based on shiftable complex directional pyramid transform and SUSAN edge detector," *Ukrainian Journal of Physical Optics*, vol. 19, no. 4, pp. 199–210, 2018.
- [16] I. Ahmad, M. Hamid, S. Yousaf, S. T. Shah, and M. O. Ahmad, "Optimizing pretrained convolutional neural networks for tomato leaf disease detection," *Complexity*, vol. 2020, Article ID 8812019, 10 pages, 2020.
- [17] E. Ayan, H. Erbay, and F. Varcin, "Crop pest classification with a genetic algorithm-based weighted ensemble of deep convolutional neural networks," *Computers and Electronics in Agriculture*, vol. 179, Article ID 105809, 2020.
- [18] X. Cheng, Y. Zhang, Y. Chen, Y. Wu, and Y. Yue, "Pest identification via deep residual learning in complex background," *Computers and Electronics in Agriculture*, vol. 141, pp. 351–356, Sep 2017.
- [19] Y. Li, H. Wang, L. Minh Dang, A. Sadeghi-Niaraki, and H. Moon, "Crop pest recognition in natural scenes using convolutional neural networks," *Computers and Electronics in Agriculture*, vol. 169, Article ID 105174, 2020.
- [20] K. Thenmozhi and U. S. Reddy, "Crop pest classification based on deep convolutional neural network and transfer learning," *Computers and Electronics in Agriculture*, vol. 164, Article ID 104906, 2019.
- [21] L. Nanni, A. Manfe, G. Maguolo, A. Lumini, and S. Brahnam, "High performing ensemble of convolutional neural networks for insect pest image detection," *Ecological Informatics*, vol. 67, Article ID 101515, 2022.
- [22] L. Chen and X. Yin, "Image recognition of typical potato diseases and insect pests using deep learning," *Fresenius Environmental Bulletin*, vol. 30, no. 8, pp. 9956–9965, 2021.
- [23] Y. Jin, "Detection of crop leaf diseases and insect pests based on improved faster R-CNN," *Fresenius Environmental Bulletin*, vol. 30, no. 6B, pp. 7278–7290, 2021.
- [24] T. Yan and X. Lv, "Combining multi-dimensional convolutional neural network (CNN) with visualization method for detection of Aphis gossypii glover infection in cotton leaves using hyperspectral imaging," *Frontiers of Plant Science*, vol. 12, Article ID 604510, 2021.
- [25] R. Jia, M. Hu, M. Zhou, et al., "An effective data augmentation strategy for CNN-based pest localization and recognition in the field," *IEEE Access*, vol. 7, pp. 160274–160283, 2019.
- [26] E. L. Mique and T. D. Palaoag, "Rice pest and disease detection using convolutional neural network," in *Proceedings of the International Conference on Information Science and System (ICISS)*, pp. 147–151, South Korea, December 2018.
- [27] P. Li, S. Yao, C. Ma, et al., "Recognition and counting of wheat mites in wheat fields by a three-step deep learning method," *Neurocomputing*, vol. 437, pp. 21–30, May 2021.
- [28] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers of Plant Science*, vol. 7, p. 1419, 2016.
- [29] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, 2018.
- [30] S.-H. Lee, S.-R. Lin, and S.-F. Chen, "Identification of tea foliar diseases and pest damage under practical field conditions using a convolutional neural network," *Plant Pathology*, vol. 69, no. 9, pp. 1731–1739, 2020.
- [31] J. Lu, J. Hu, G. N. Zhao, F. H. Mei, and C. S. Zhang, "An in-field automatic wheat disease diagnosis system," *Computers and Electronics in Agriculture*, vol. 142, pp. 369–379, 2017.
- [32] J. Ubbens, M. Cieslak, P. Prusinkiewicz, and I. Stavness, "The use of plant models in deep learning: an application to leaf counting in rosette plants," *Plant Methods*, vol. 14, no. Jan, p. 6, 2018.