A Novel Crop Weed Recognition Method Based on Transfer Learning from VGG16 Implemented by Keras

Zichao Jiang
Indian Springs School, Alabama, USA
Hzhe_wjdi@163.com

Abstract. This study used a method based on convolutional neural network model, VGG16, to identify images of weeds in the field. As the basic network, VGG16 has very good classification performance, and the network structure is unconventional. It is relatively easy to modify. It can fine-tune other data sets on this basis. Therefore, the transfer learning method is applied to our own Kaggle competition website. Download the weed data set. The site covers approximately 3,500 images in 12 categories. Due to limited data and computational power, our model fixes the first 14 layers of VGG16 parameters for layer-by-layer automatic extraction of features, adding an average pooling layer, convolution layer, Dropout layer, fully connected layer, and softmax for classifiers. The layer has a total of 5 layers, for a total of 19 layers. The experimental results show that the final model performs well in the classification effect of 12 weed images. The accuracy rate on the training set is 98.99%, and the accuracy on the verification set is 91.08%. It can be applied to crop weed identification. It provides accurate and reliable judgment basis for positioning and quantitative chemical pesticide spraying, and is the key to achieving refined agriculture.

1. Introduction
Herbicides are chemicals commonly used in agriculture to remove unwanted plants. While preventing unwanted plants from competing for nutrients with crops, the use of herbicides can also cause many health and environmental side effects, as well as weed resistance to herbicides. With the latest developments in machine learning, models can be trained to determine whether plants are crops or weeds for precise removal.

Through relevant research, the author finds that in the identification of crop weeds, not only the related practical technology research is relatively small, but also the manual related features based on agricultural knowledge are used to identify weed types, and there are incomplete design features and identification of weeds. Problems such as limited type and low recognition accuracy. With the rise of deep neural networks, especially in the field of computer vision, various models for identifying image classifications on ImageNet datasets, especially with the 2012 classic convolutional neural network AlexNet in the field of image classification A huge breakthrough, various advanced classification image algorithms have appeared, such as four classic networks: AlexNet, VGGNet, Google Inception Net, ResNet, as shown below:
The four networks are arranged in the order in which they appear, and the depth and complexity are also progressive. They won the 2012 Champion of the ILSVRC (ImageNet Large Scale Visual Recognition Challenge) competition classification project (top-5 error rate 16.4%, using additional data to reach 15.3%, 8-layer neural network), 2014 runner-up (top-5) The error rate is 7.3%, 19-layer neural network), the 2014 champion (top-5 error rate 6.7%, 22-layer neural network) and the 2015 champion (top-5 error rate 3.57%, 152-layer neural network).

In summary, this study intends to combine the advanced convolutional neural network structure to identify field weeds and achieve certain breakthroughs and innovations in the identification technology of crop weeds.

2. Method

2.1. Neural Network

The most widely used definition is the description of Kohonen in 1988. The neural network is a network of extensive parallel interconnects consisting of adaptive simple units whose organization can simulate the interaction of biological nervous systems with real-world objects. Reaction.

The neural network is mainly composed of a neuron model, a threshold, an activation function, and the like. A neuron model, a simple unit in the above definition.

In a biological neural network, each neuron is connected to other neurons. When it is excited, it sends chemicals to the connected neurons, thereby changing the potential within these neurons; if a neuron has a potential that exceeds one Threshold, then it will activate, that is, excited. Send chemicals to other neurons. We abstract this neuron model above, which is the M-P neuron model:

![M-P neuron Model](image1)

The threshold, also known as bias, has a similar meaning to the valve. In the calculation, the threshold can be regarded as the connection weight corresponding to a dummy node with a fixed input of -1.
The ideal activation function is a step function, but in practice, because the step function is not smooth and discontinuous, we generally use sigmoid or tanh functions, as follows:

![Sigmoid activation function](image)

Figure 3. Sigmoid activation function

2.2. Convolutional neural network

The Convolutional Neural Network (CNN) consists of repeated convolutional and pooled layers and is widely used for image recognition. The convolution layer slides multiple filters across the image to generate a matrix, and the pooling layer samples from the convolutional layer. CNN is especially useful when extracting features from images because it preserves spatial information. The output of the convolution and pooling layers feeds into a dense neural network, producing a final output.

Convolutional neural networks usually contain the following layers:

- Convolutional layer, each convolutional layer in a convolutional neural network is composed of several convolutional units, and the parameters of each convolution unit are optimized by a back propagation algorithm. The purpose of the convolution operation is to extract different features of the input. The first layer of convolutional layer may only extract some low-level features such as edges, lines and corners. More layers of networks can iteratively extract more complex from low-level features. Feature.

- Rectified Linear Units layer (ReLU layer), this layer of nerve activation function (Rectified Linear Units, ReLU).

- The pooling layer usually obtains a feature with a large dimension after the convolution layer. The feature is cut into several regions, and the maximum or average value is obtained to obtain a new feature with a small dimension.

- The Fully-Connected layer combines all local features into global features to calculate the score for each of the last classes.

An application example of each layer of a convolutional neural network:

![Convolutional neural network for classification of car image recognition](image)

Figure 4. Convolutional neural network for classification of car image recognition
2.3. **VGG16 Network**

Fig. 5. are separately structure of AlexNet and VGG16, (a) is AlexNet, (b) is VGG16, as follows:

![AlexNet Network](image1)

![VGG16 Network](image2)

**Figure 5.** Comparison of AlexNet and VGG16 Network

One improvement of the VGG16 over AlexNet is the use of several consecutive 3x3 convolution kernels instead of the larger convolution kernels in AlexNet (11x11, 7x7, and 5x5). For a given receptive field (the local size of the input picture associated with the output), the use of stacked small convolution kernels is superior to the use of large convolution kernels, because multiple layers of nonlinear layers can increase network depth to ensure more complex learning. The pattern, and the cost is still relatively small (less parameters). In simple terms, in VGG, three 3x3 convolution kernels are used instead of 7x7 convolution kernels, and two 3x3 convolution kernels are used instead of 5x5 convolution kernels. The main purpose of this is to ensure the same under the condition of sensing the wild, the depth of the network is improved, and the effect of the neural network is improved to some extent.

2.4. **Transfer Learning**

With the development of computing hardware and algorithms, the lack of tagged data has become more prominent. Not every field will spend a lot of manual annotations like imagenet to produce some data, especially for the industry, which is produced every moment. A large amount of new data, labeling these data is a time-consuming and laborious task. Therefore, although supervised learning can solve many important problems, it has certain limitations. Based on such an environment, migration learning has changed. Particularly important.
As the basic network, the VGG16 model is a successful application of convolutional neural networks in the field of image classification algorithms. Its network structure is unconventional and easy to modify. The model trained on ImageNet has been published, and other data sets can be fine-tuning on this basis. And adaptability to other data sets is very good, so it is easy to carry out the migration learning application to the tagged data set under the specific problem scenario.

2.5. Data Augmentation

Our weed dataset comes from the kaggle competition website, a total of 12 categories of 3,500 images, of which 80% of the data is randomly divided into training sets, 20% of the data is divided into verification sets, so the training set and verification set obey the statistical sense The same distribution, can better test the effect of our model, the data set is as follows:

![Figure 6. 12 types of weed data sets](image)

In the early days of the experiment, we found that the classification of the model is not good. After analysis, because the picture of each type of weed contains pictures of different periods of its growth, this is difficult for the network. After finishing the picture, the network effect should be very good. Ok, so we finally use the cull training set and the verification set with a resolution less than 300*300, and get 1784 12 kinds of weed pictures for the training set, 457 12 kinds of weed pictures for the verification set, as shown below Shown as follows:

![Figure 7. Summary of data after pretreatment of 12 types of weeds](image)

In view of the total number of 1784+457=2241 pictures on the training set and the verification set after data preprocessing, the lack of data may lead to the under-fitting problem of our model structure, that is, the convolutional neural network model is on the training set and the test set. The performance is poor, there is no good feature of extracting weed pictures on the training set. We use the data
enhancement method common in deep learning to input the image-enhanced data set into our model for training and verification.

The role of data enhancement: 1. Increase the amount of training data, improve the generalization ability of the model, 2. Increase the noise data, and improve the robustness of the model. The commonly used data enhancement methods are mainly realized by the rotation, translation and deformation of the image.

The image rotation is integrated in our weed dataset, and the original image is converted into multiple different images by setting the zoom ratio to achieve image scaling and panning. Because convolutional neural networks remain invariant to translation, viewing angle, size, or illuminance (or combinations of the above), specifically: a convolutional neural network has a property called invariance, even if the convolutional neural network is placed differently in the direction, it can also classify objects. As the picture shows:

(a) Raw image               (b). Three images of processing by rotation, scaling, zooming

Figure 8. Image data augmentation

3. Experiment Design
In the experimental design process, we did not fully adopt the VGG total 19-layer structure model. For our experimental data 12-type weed dataset, we reserved the pre-training model from the input layer to the block4_conv3 layer, in which the input layer input image The size is (224, 224, 3) and we need to convert the image in the dataset to the target input size. The global average pooling layer, two fully connected layers, and one Dropout inactive layer are sequentially added (the main function is to avoid over-fitting of the model during the training process, parameter setting .2, indicating 128 in the second full connection) Neurons were randomly inactivated by 20%), and the last classifier using softmax (for the probabilistic output of categories for 12 types of weed data). The model structure and parameters are as follows:

```
global_average_pooling2d_1 (None, 512) 0
dense_1 (Dense) (None, 512) 262656
dense_2 (Dense) (None, 128) 65664
dropout_1 (Dropout) (None, 128) 0
dense_3 (Dense) (None, 12) 1548
```

Figure 9. Our model fine-tuning based on VGG16 pre-training

The model has a total of 19 layers. As shown in the above figure, we can fix the first 14 layers of VGG16 parameters, namely 7,635,264, because of the limitations of the data set and the limitations of the computing power. This can greatly reduce the model training time and quickly. Try different model structure ideas, only training the data set of our last added 5-layer model, the total number of parameters that need to be trained is 329,868.
4. Experiment Process
After the data is organized, we begin our experimental process. First we set our model hyperparameters, epoches we set 500, Batch_size is set to 20, fixed batch size, set the number of training set steps to the total number of training sets in the sample divided by the batch size, each round completes all layers The model weight and all the deviation parameters are updated. The number of steps on the verification set is the total number of verification sets in the sample divided by the batch size, which is batch_size(20). At the same time, we calculate the loss function value in the back propagation process. rmsProp(root mean square prop) optimizer, and set the learning rate to 0.001, the learning rate is attenuated to 0.0001, and the reasonable setting of the learning rate can reduce the gradient of the loss function between the model output and the expected value, which can ensure reasonable training. Time gets the convergence of the local loss function, and can avoid the learning rate setting too large. Here, the learning rate attenuation method is used to make the model miss the best local loss value, and the parameters in the model structure such as weights and deviations are back-propagated. A reasonable update is obtained in the algorithm. The specific model training process is shown in the following figure:

5. Data Analysis
On a workstation with a GPU (Geforce GTX 1080, 8GB memory) accelerator, we ran for 500 rounds and used the matplotlib that comes with Python to draw the results of the epoches and loss, epoches and acc in the train set and the validation set respectively. As follows:

Figure 10. Network training process by keras

Figure 11. Experimental result of 500 epochs
It can be seen from Fig.11. That after 500 rounds of training on the model, the network structure based on VGG16 migration learning reaches 0.9899 on the training set and 0.91 on the verification set, indicating that our model performs well and can be used for 12 types of weed data. The classification and recognition are very good, and the classification effect is good. The reliability of the image recognition classification based on the VGG16 migration learning method is basically verified. It also shows that the deep neural network, especially the convolutional neural network, has indeed achieved a large image field. Success provides a good tool and method for our weed identification applications.

6. Conclusion and Discussion

This study is a weed dataset on the public data source Kaggle website. Based on the convolutional neural network (VGG16) migration learning, we can identify 12 types of weed images in different growth stages with high precision and achieve the intended purpose. However, there is still much room for improvement in practical applications. For example, in the actual crop removal of weeds, the weed category is far more than 12 types. If we can collect more different types of weed pictures, we can not only improve the model. The classification effect also has greater practical value. In addition, the neural network model is only a tool and an algorithm. It needs to be combined with intelligent hardware such as a drone to automatically realize the positioning and quantitative spraying of intelligent chemical agents on the basis of identification. This can not only realize the economic significance of increasing crop yield and income, but also protect it. A good ecological environment allows people to coexist harmoniously with nature and realize the need for sustainable agricultural development in technology.

References

[1] Feijian Jin. Research on feature extraction and recognition of weed image based on machine vision [D]. Jiangsu University,2007.
[2] Xiaoxia Feng. Research on image recognition algorithm based on deep learning [D]. Taiyuan University of Technology, 2015.
[3] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]. Advances in neural information processing systems. 2012: 1097-1105.
[4] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1-9.
[5] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv: 1409.1556, 2014.
[6] P. Ekman, He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.
[7] Hinton, G.E., “Deep Belief Nets,” 2017 NIPS Tutorials Notes, Neural Information Processing Systems Conference, Vancouver, BC, December 2007.
[8] Bengio, Y. and Lecun, Y., “Scaling Learning Algorithms Towards AI,” In Large Scale Kernel Machines, 2007a.
[9] Delalleau, O. and Bengio, Y., “Shallow vs. Deep Sum-Product Networks,” In NIPS, 2011.
[10] LeCun, Y., Jackel, L.D., Boser, B., Denker, J.S., Graf, H.P., Guyon, I., Henderson, D., Howard, R.E., and Hubbard, W., “Handwritten Digit Recognition: Application of Neural Network Chips and Automatic Learning,” IEEE Communication Magazine, 27 (11), 41-46, 1989.
[11] Chen, B., Ting, J. A., Marlin, B. M., and de Freitas, N., “Deep Learning of Invariant Spatial-Temporal Features from Video,” NIPS*2010 Deep Learning and Unsupervised Feature Learning Workshop, 2010.
[12] Kingma,D. and Ba,J., “Adam: A Method for Stochastic Optimization,” arXiv preprint arXiv: 1412.6980, 2014