Computers and Electronics in Agriculture 141 (2017) 351–356

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Pest identification via deep residual learning in complex background

Xi Cheng, Youhua Zhang, Yiqiong Chen, Yunzhi Wu, Yi Yue *

College of Information and Computer, Anhui Agricultural University, Hefei, China

ARTICLE INFO

Article history:
Received 15 April 2017
Received in revised form 4 August 2017
Accepted 5 August 2017

Keywords:
Agricultural pests
Image recognition
Convolutional neural networks
Deep residual learning
ResNet

ABSTRACT

Agricultural pests severely affect both agricultural production and the storage of crops. To prevent damage caused by agricultural pests, the pest category needs to be correctly identified and targeted control measures need to be taken; therefore, it is important to develop an agricultural pest identification system based on computer vision technology. To achieve pest identification with the complex farmland background, a pest identification method is proposed that uses deep residual learning. Compared to support vector machine and traditional BP neural networks, the pest image recognition accuracy of this method is noticeably improved in the complex farmland background. Furthermore, in comparison to plain deep convolutional neural networks such as Alexnet, the recognition performance in this method was further improved after optimized by deep residual learning. A classification accuracy of 98.67% for 10 classes of crop pest images with complex farmland background was achieved. Accordingly, the method has a high value of practical application, and can be integrated with currently used agricultural networking systems into actual agricultural pest control tasks.

1. Introduction

Agricultural pests have long posed a severe threat to the growth of crops and the storage of agricultural products. Every year, agricultural pests cause significant losses on a global scale. Just relying on low-speed and inefficient manual identification will not meet the actual needs, and causes high labor costs. In recent years, agricultural networking is developing rapidly and through the use of cameras in the farmland, images of agricultural pests can be more convenient accessed. Computer vision based image recognition technology can effectively reduce the recognition cost, and both recognition speed and efficiency can be significantly improved. However, in contrast to sample photos, the actual images obtained from the farmland often have high background noise. If feature pre-processing is not conducted, the traditional machine learning classification methods such as support vector machine and back propagation (BP) neural network often cannot achieve satisfactory accuracy rates. The farmland environment is complex, thus it is often difficult to select general features that are suitable for all target pests. However, the above methods are more influential in the selection of characteristics, so it is often difficult to integrate the actual environment. During the past two years, the deep learning technology has been developing rapidly. Deep convolutional neural networks (CNN) have been applied to the field of image recognition, such as street view recognition (Goodfellow et al., 2013), vehicle detection (Chen et al., 2014), human motion recognition (Ji et al., 2013), and further applied to audio and video recognition (Abdel-Hamid et al., 2014; Yue-Hei Ng et al., 2015), where it achieved very good results. The CNN have the ability to automatic extract image features, and can thus be used as a general feature extraction tool and applied to agricultural pest identification in the farmland environment. To prevent degradation of deep neural networks and improve training of the deeper convolutional neural network models, it is necessary to use deep residual learning. Therefore, this research utilizes the CNN and residual learning to build a system for identifying agricultural pest images taken from actual farmland with complex background. The system is sufficiently robust to recognize pests with assimilatory colouration and could be trained end to end which is more practical for agricultural tasks than previous researches.

2. Related works

Research on identification of agricultural pests based on computer vision has been a hot topic. In recent years, many pest recognition systems were proposed. Larios et al. (2008) proposed a SIFT-based feature learning method and constructed a feature histogram to classify stone fly larvae images. Zhao et al. (2009) studied image recognition of pests of sugarcane cotton aphids based on rough set and fuzzy C-means clustering. Zhu and Zhang (2010)

* Corresponding author.
E-mail addresses: opteroncx@ahau.edu.cn (X. Cheng), yyyue@ahau.edu.cn (Y. Yue).

http://dx.doi.org/10.1016/j.compag.2017.08.005
0168-1699/

established an insect automatic classification system by analyzing both the color histogram and the gray level co-occurrence Matrices of the insect wing. Faithpraise et al. (2013) proposed a plant pest identification system based on k-means cluster and correspondence filters. Chengjun et al. (2016) used a spatial pyramid with sparse coding to identify farmland pest images. Compared to the early support vector machine and the neural network methods, the recognition accuracy of pest images with background has been improved. To further enhance the insect recognition ability, Xie et al. (2015) developed an insect recognition method based on multi-task sparse representation and multi kernel learning techniques.

However, almost all of the above methods require a complex pretreatment of pest images, and the performance of the model is often influenced by characteristics of the selected features. Most of the pest image samples used in earlier studies are images with a uniform background or require removal of the background or binarization. Through the convolutional neural networks, end-to-end training of pest images with background can be achieved, thus greatly simplifying the training process.

3. Pest image recognition

3.1. Convolutional neural networks

Neural networks have been proven to be good classifiers for those linear inseparable problems and many developments were achieved on the structure of networks to enhance the performance of classification or clustering (Zhao and Huang, 2007; Zhao et al., 2010). When dealing with image classification problems, the most advanced models are convolutional neural networks. The history of convolutional neural networks originated from Hubel and Wiesel (2009) and where developed for the study of animal visual cortical cells; in 1980, Fukushima (1980) proposed a model named Neocognitron, which can be considered as the embryonic form of convolutional neural networks. In the late 1990s, Lecun et al. (1989), LeCun et al. (1998) proposed a modern structure of convolutional neural networks. Constrained by computer performance, deeper neural networks were difficult to train; therefore, early neural networks often only contained one hidden layer. In recent years, the rapid increase of computer performance and the introduction of GPU aided computing (Coates et al., 2009) enables the construction and training of deep neural network models.

In 2012, Krizhevsky et al. (2012) presented the Alexnet, a deep convolutional neural network model, which won the results in an ILSVRC2012 image recognition competition due to significantly outperforming other methods. Then in 2014, VGG (Simonyan and Zisserman, 2014) and GoogLeNet (Szegedy et al., 2015) were proposed, further enhancing recognition performance. To effectively train a very deep network structure, He et al. (2016) proposed the concept of deep residual learning, developing ultra-high depth like ResNet-152 (He et al., 2016) and Inception-ResNet-v2 (Szegedy et al., 2016).

3.2. Structure of CNN

The basic structure of the convolutional neural networks is shown in Fig. 1, a basic convolutional neural network includes the input layer, convolution layer, pooling layer, full connected layer, Softmax classification, and output layer.

The convolution layer can be represented with the following formula. During a convolution process of a certain layer, a filter slides over that layer and its weight matrix does Hadamard product with the values of the pixels below the filter.

$$x_j^l = f\left(\sum_{i\in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right) \tag{1}$$

In the case where b_j^l represents a bias term and f (.) represents an activation function, the general neural network often uses the Sigmoid function or the hyperbolic tangent function with a range of [−1, 1]. However, with the increasing network depth, gradients are often prone to vanish or explode, to suppress the above phenomenon; therefore, the ReLu (Krizhevsky et al., 2012) activation function was proposed.

The working process of the pooling layer is down-sampling, mainly including the maximum pooling and average pooling methods (Boureau et al., 2010). The process can be expressed with the following formula.

$$x_j^l = \mathrm{down}(x_i^{l-1}) \tag{2}$$

In the process of pooling, a window of specified size is also allowed to slide on a feature map of a certain layer. If the maximum pooling was used, the maximum value in the window would be retained. If the average pooling was used, the average value in the window would be retained.

The top of the convolutional neural networks that deals with the multi-classification problem often used the Softmax classifier. In Softmax regression, the probability that an input x belongs to class t can be expressed by the following equation:

$$P(y_i = t|x_i;\theta) = \frac{e^{\theta_t^T x_i}}{\sum_{l=1}^T e^{\theta_l^T x_i}} \tag{3}$$

The loss function is as follows: It is the cross entropy loss between the output probability vector and the actual class vector.

$$\mathrm{J}(\theta) = -\frac{1}{N}\left[\sum_{i=1}^N \sum_{i=c}^C 1\{y_i = t\}\log\frac{e^{\theta_t^T x_i}}{\sum_{l=1}^T e^{\theta_l^T x_i}}\right] \tag{4}$$

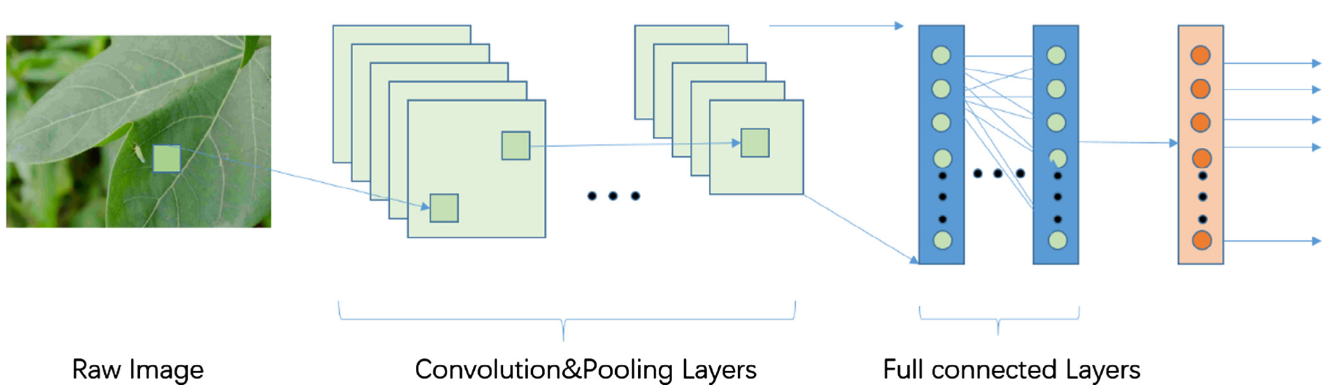

Fig. 1. Basic structure of CNN.