Computers and Electronics in Agriculture 153 (2018) 46–53

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification

Jayme Garcia Arnal Barbedo

Embrapa Agricultural Informatics, Av. André Tosello, 209 - C.P. 6041, Campinas, SP 13083-886, Brazil

ARTICLE INFO

Keywords:
Image processing
Deep neural nets
Image database
Disease classification

ABSTRACT

The problem of automatic recognition of plant diseases has been historically based on conventional machine learning techniques such as Support Vector Machines, Multilayer Perceptron Neural Networks and Decision Trees. However, the prevailing approach has shifted to the application of deep learning concepts, with focus on Convolutional Neural Networks (CNNs). In general, this kind of technique requires large datasets containing a wide variety of conditions to work properly. This is an important limitation, given the many challenges involved in the construction of a suitable image database. In this context, this study investigates how the size and variety of the datasets impact the effectiveness of deep learning techniques applied to plant pathology. This investigation was based on an image database containing 12 plant species, each presenting very different characteristics in terms of number of samples, number of diseases and variety of conditions. Experimental results indicate that while the technical constraints linked to automatic plant disease classification have been largely overcome, the use of limited image datasets for training brings many undesirable consequences that still prevent the effective dissemination of this type of technology.

1. Introduction

The image-based classification of plant diseases is a difficult problem with a wide variety of challenges associated, including the presence of symptoms with extensive range of visual characteristics, possibility of multiple simultaneous disorders in a single plant, and different disorders having similar symptoms, among others (Barbedo, 2016). Extrinsic factors such as interference caused by the image background and illumination variations associated to capture conditions add even more complexity to the problem. While the combination of image processing and machine learning has led to many advances (Barbedo, 2013), practical use of tools like these has been limited. In the last few years, several studies have used the concepts of deep learning, and Convolutional Neural Networks (CNN) in particular, to try and make this kind of tool more accurate (Table 1).

Deep learning is a branch of machine learning composed by a number of algorithms that try to model high-level data abstractions using a deep graph with several processing layers containing linear and non-linear transformations (Goodfellow et al., 2016). Because CNNs have an intimate relationship between layers and spatial information, they are well-suited for image classification tasks (Arel et al., 2010), which explains their prevalence in recent plant disease classifiers. This type of neural network usually requires a very large number of samples for proper training, but this constraint can be relaxed by the application of transfer learning. This technique recycles previously trained networks by using the new data to update a small part of the original weights (Bengio, 2012).

Many of the studies found in the literature use transfer learning in their experiments (Mohanty et al., 2016; Brahimi et al., 2017; Ferentinos, 2018; Liu et al., 2018), and those that do not apply this technique use CNN architectures that are similar to existing ones (Amara et al., 2017; DeChant et al., 2017; Lu et al., 2017; Oppenheim and Shani, 2017). Also, many studies employed the initial PlantVillage dataset (Mohanty et al., 2016; Brahimi et al., 2017), which contains images that were mostly collected using a regularized process that generated relatively homogeneous backgrounds (Hughes and Salathé, 2015; Mohanty et al., 2016).

Thus, many studies are applying similar tools to a dataset that does not reproduce the range of conditions expected to be found in practice. This explains why most results reported in the literature show nearly perfect accuracy, without much variation between studies. It is quite revealing that when Mohanty et al. (2016) applied the model trained using the PlantVillage database to images originated from trusted online sources, the accuracy quickly fell below 50%. On the other hand, some studies applied their own datasets, but those were either collected under controlled conditions (Liu et al., 2018), and/or include only a few

E-mail address: jayme.barbedo@embrapa.br.

https://doi.org/10.1016/j.compag.2018.08.013
Received 30 March 2018; Received in revised form 10 July 2018; Accepted 5 August 2018
0168-1699/

Table 1
Studies employing deep learning for plant disease recognition. The accuracy is given by the number of samples correctly classified divided by the total number of samples.

Reference	CNN Network	Dataset	Accuracy	# Classes
Amara et al. (2017)	LeNet architecture	PlantVillage (extended)	92–99%	3
Brahimi et al. (2017)	AlexNet, GoogLeNet	PlantVillage	99%	9
Cruz et al. (2017)	Modified LeNet	Olive tree images (own)	99%	3
DeChant et al. (2017)	Pipeline	Corn images (own)	97%	2
Ferentinos (2018)	Several	PlantVillage (extended)	99%	58[a]
Fuentes et al. (2017)	Several	Tomato images (own)	83%	10
Liu et al. (2018)	AlexNet	Apple images (own)	98%	4
Lu et al. (2017)	AlexNet inspired	Rice images (own)	95%	10
Mohanty et al., 2016	AlexNet, GoogLeNet	PlantVillage	99%	38[b]
Oppenheim and Shani (2017)	VGG	Potato images (own)	96%	5

[a] Classes are distributed among 25 plant species.
[b] Classes are distributed among 14 plant species.

classes (Dechant et al., 2017; Fuentes et al., 2017; Lu et al., 2017). While all these studies yielded important contributions to the field, dataset limitations still prevent broader practical use.

This situation is in large part caused by the difficulties involved in building truly comprehensive databases. Most relevant visual manifestations of diseases happen in the field, as experiments with controlled inoculations often cannot produce the symptom variety found under more realistic conditions. Additionally, the visual characteristics of a symptom may change as the disease progresses and environmental factors such as humidity and temperature oscillate, so pictures may have to be taken frequently in order to cover the entire range of possibilities. It is also important to consider that all images need to be labeled with the correct disease, which is often a labor-intensive and error-prone process (Barbedo, 2018).

These circumstances require a better understanding about the effects of using relatively small datasets on the effectiveness of deep learning tools for plant disease classification. This is the objective and main contribution of this study. An image database, containing 12 plant species with very distinct characteristics in terms of number of samples and diseases, was used to test the behavior of CNN under a variety of conditions. The insights drawn from the experimental results led to a better understanding about the strengths and limitations of deep learning networks when these are trained with datasets of limited size and diversity. As a result, it was possible to draw some conclusions about the current development of deep learning-based plant disease classifiers, as well as to suggest some potential targets for future research on the subject. The database used in this work is being made freely available for academic purposes at a repository in the address https://www.digipathos-rep.cnptia.embrapa.br/.

2. Material and methods

2.1. Image dataset

The database available in the repository includes images of symptoms expressed not only on leaves, but also on stems, flowers and fruits. In this investigation, only images containing leaves were used in order to make the data more consistent. As a result, the image dataset used in this work is similar to the one used in Barbedo (2016). However, some diseases were removed from the original ensemble as they had too few images to be properly handled by CNNs, resulting in 56 diseases infecting 12 plant species. Since this dataset has already been detailed in Barbedo (2016), only a brief description is presented here. Additionally, only the common names of plants and diseases are presented; scientific names can be found in the database repository.

Table 2 shows how the database is distributed in terms of plant species and disorders. Images were captured using a variety of digital cameras and mobile devices, with resolutions ranging from 1 to 24 MPixels. About 15% of the images were captured under controlled conditions, and the remainder 85% of the images were captured under real conditions, with the leaves attached to the host plant. All images were stored in the 8-bit RGB format.

2.2. Experimental setup

Transfer learning (Bengio, 2012) was applied to a pretrained CNN (GoogLeNet) using the Neural Network Toolbox available in Matlab 2017b. The GoogLeNet architecture was chosen because of its superior performance in the context of plant disease recognition (Mohanty et al., 2016; Ferentinos, 2018). The parameters used to train the network were the following: Base Learning Rate, 0.001; Momentum, 0.9; Mini Batch Size, 16; Number of Epochs, 5. All experiments were run using a NVIDIA Quadro K620 Graphics Processing Unit (GPU).

In order to investigate the influence of the background on the results, two separate CNNs were retrained, the first using the original unprocessed images, and the second using whole images with background manually removed. In each case, 80% of the samples were used for training and 20% for validation. All images were resized prior to training to meet GoogLeNet's input dimension requirement ($224 \times 224 \times 3$ pixels).

In order to increase the size of the training set and decrease overfitting problems (Liu et al., 2018), the training datasets were augmented using a number of operations (Fig. 1).

The results are presented as confusion matrices with an overall accuracy associated (Table 3). The confusion matrices are given in terms of percentages, not absolute numbers. Those values were obtained using a 10-fold cross-validation. It is important to remark that the number of images, diseases and conditions for each plant species varies significantly. This allowed an investigation on the performance of the CNN under a wide range of different conditions and contexts.

3. Results

Table 3 presents the overall accuracies obtained for each plant species, considering the original and background removed images. Because background removal was explicitly investigated with separate CNNs, this factor was used to organize this section, following the four different behaviors that were observed: (a) no significant impact on the accuracies; (b) substantial accuracy improvement; (c) substantial accuracy decrease; (d) mixed results. Each subsection contains one pair of confusion matrices obtained for a selected plant species. The confusion matrices obtained for the other plant species are omitted due to space constraints.

3.1. Small background removal impact

Crops for which the impact of background removal was mild had in common the characteristic of having few classes (up to four) with