IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 22, 2021, accepted November 19, 2021, date of publication November 25, 2021,
date of current version December 9, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3131002

Convolutional Neural Networks for Texture Feature Extraction. Applications to Leaf Disease Classification in Precision Agriculture

STEFANIA BARBURICEANU, SERBAN MEZA, (Member, IEEE), BOGDAN ORZA, RAUL MALUTAN, AND ROMULUS TEREBES, (Member, IEEE)
Communications Department, Technical University of Cluj-Napoca, 400114 Cluj-Napoca, Romania

Corresponding author: Stefania Barburiceanu (stefania.barburiceanu@com.utcluj.ro)

This work was supported by the Project "Entrepreneurial Competences and Excellence Research in Doctoral and Postdoctoral Programs—ANTREDOC," Project co-funded by the European Social Fund Financing under Grant 56437/24.07.2019.

ABSTRACT This paper studies the use of deep-learning models (AlexNet, VggNet, ResNet) pre-trained on object categories (ImageNet) in applied texture classification problems such as plant disease detection tasks. Research related to precision agriculture is of high relevance due to its potential economic impact on agricultural productivity and quality. Within this context, we propose a deep learning-based feature extraction method for the identification of plant species and the classification of plant leaf diseases. We focus on results relevant to real-time processing scenarios that can be easily transferred to manned/unmanned agricultural smart machinery (e.g. tractors, drones, robots, IoT smart sensor networks, etc.) by reconsidering the common processing pipeline. In our approach, texture features are extracted from different layers of pre-trained Convolutional Neural Network models and are later applied to a machine-learning classifier. For the experimental evaluation, we used publicly available datasets consisting of RGB textured images and datasets containing images of healthy and non-healthy plant leaves of different species. We compared our method to feature vectors derived from traditional handcrafted feature extraction descriptors computed for the same images and end-to-end deep-learning approaches. The proposed method proves to be significantly more efficient in terms of processing times and discriminative power, being able to surpass traditional and end-to-end CNN-based methods and provide a solution also to the problem of the reduced datasets available for precision agriculture.

INDEX TERMS Applied convolutional neural networks, leaf disease detection, image classification, texture classification, texture feature extraction.

I. INTRODUCTION

Image feature extraction and classification is a computer vision field that has been studied intensively by researchers due to its practical relevance for various scenarios, including that of precision agriculture, [1]. Plant diseases have a huge effect on the agricultural productivity [2]. They can easily degrade the quality of the products, so they must be detected as soon as possible. The current methodology for detection is the human perception of plant leaves [3]. However, this method is not efficient in terms of available resources, especially for large crops, and automatic image classification systems can be beneficial in this situation. In the literature, several plant disease classification problems were addressed, such as the classification of cucumber and citrus leaves [4], [5] which is performed by using the Gray-Level Co-occurrence Matrix (GLCM) for the extraction of relevant features. In [6], colour information is used along with GLCM - derived features and Gabor characteristics for the classification of mango leaves. Deep-learning methods are also mentioned for the classification of plant diseases in [7]–[9].

Until recently, [7], [9], the problem of image classification has been addressed as a two-stage approach: the extraction of handcrafted features and machine-learning classification. The feature extraction step is regarded as the most important stage because the subsequent classification task is based

The associate editor coordinating the review of this manuscript and approving it for publication was Gangyi Jiang.

on the derived image descriptors. Even the most powerful machine-learning classifier will provide a poor classification performance if the image features are not chosen appropriately. The extraction of relevant and discriminative features is a challenging task for real-world applications. Moreover, images are captured under various conditions and, to obtain good classification results, the extracted features should provide invariance to several transformations (such as scale, rotation, illumination conditions) and robustness to noise.

One of the most popular and efficient feature extraction methods is the Local Binary Patterns operator (LBP) [10] and its improved version, proposed in [11]. The LBP descriptor is based on the signs of differences between neighbouring pixels, and it is used to describe locally the texture of the analysed image. Later, several LBP-derived operators which provide improved invariance to different transformations and a greater discrimination power were proposed, such as the Median Robust Extended Local Binary Patterns (MRELBP) [12]. Also, in order to improve the robustness to Gaussian noise, the Block Matching and 3D Filtering Extended Local Binary Patterns (BM3DELBP) was introduced by us in [13]. Another popular texture feature descriptor is the Gray-Level Co-occurrence Matrix (GLCM) [14] which achieved significant performance for texture classification tasks as reported in the literature.

In the case of traditional machine-learning methods, expert-driven feature selection and extraction are needed. A specialist must design a feature extraction method capable of outputting the most relevant features and feed them into a conventional machine-learning classifier. The classifier is then trained to learn from data and apply the learnt information to new data in order to make a classification decision.

However, lately, [7], [9], impressive results were obtained with the use of deep-learning methods, revolutionising the image and object classification field. Rather than relying on handcrafted features, these methods can be used as end-to-end approaches because they work by automatically learning the relevant features themselves, without the need of expertise, from the raw data provided as input. Deep-learning methods are constructed to learn hierarchically, their architecture being composed of several hidden layers, and are generally trained on large datasets to obtain a good classification performance. Such a dataset is ImageNet [15]. The main disadvantage of these algorithms is the long training time, which in most situations is a lot larger compared to the case of traditional classification methods. This is due to the large number of parameters that have to be learnt from the data.

The Convolutional Neural Network (CNN) is a deep-learning technique that has been widely used in the past years with great success for many computer vision tasks, [1], [7], [9], [16]. The architecture of a CNN is composed of several types of layers: convolutional, nonlinear, pooling, fully connected, normalization and others. The stack of convolutional, nonlinear, and pooling layers act as a feature extractor. The second part of the CNN is composed of several fully connected layers that are used to make a classification decision based on the generated features. We show in Fig. 1 the general block scheme of an end-to-end CNN architecture for classification tasks.

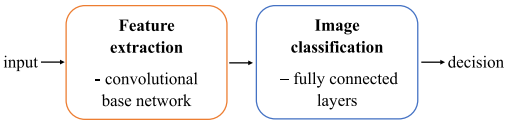

FIGURE 1. The architecture of an end-to-end CNN for image classification.

One of the main disadvantages of CNN-based methods is the fact that very large datasets are required in order to achieve significant results, [17], like with any deep-learning technique. However, there are applications in which the number of available training samples is limited, [18], especially because of the large resources (time, expertise, etc.) needed to acquire and label consistently a vast number of images (e.g. precision agriculture). This is largely addressed either by performing some sort of ''data augmentation,'' where, from the existing data, ''new'' data is generated, or by deploying what is termed ''transfer learning.''

Data augmentation is a challenging approach, as it tries to create relevant variability in the data, and, with the use of generative adversarial networks contributes more to the increase in the overall complexity of the classification system. The work in [17] provides a relevant overview of the field, and [19] is an example of an applied case of vine leaf classification.

The ''transfer learning'' concept, developed by [20], [21], resembles the approaches we, as humans, take in our everyday life, as we do not learn everything from scratch, but rather use the knowledge gained in a particular previous task in other related new tasks. Practically, we transfer the knowledge acquired in the past to solve future problems. Isolated training models are designed specifically for a particular task and dataset, whereas in the transfer learning models, the gained knowledge can be transferred and used in another related new task which can imply a better performance obtained on a smaller dataset and less training time. CNN-based methods that explored the transfer learning approach by using features derived from pre-trained CNNs on large image datasets can be found in the work of [22]–[26] and others.

Typically, a new object classification problem is addressed by using a pre-trained model without its classification layers to extract the relevant features for the new problem. Practically, the weights of the network are not updated for the new task, but they are used in the new problem exactly as they were trained for the previous task and only the classification part is replaced. Popular CNN models and datasets which are widely used for feature extraction in the context of transfer learning and belonging to the object classification problem include: AlexNet [27], VggNet [28], GoogleNet [29], and ResNet [30]. Features can be extracted either from the convolutional layers or from the fully connected layers of the network. In general, it was shown, [23], [31], [32], that the

features extracted from convolutional layers have a better generalization capability. The features extracted from fully connected layers have a poorer transferability because they are more specific to a particular task or dataset.

In the context of plant disease, the classification problem translates to a texture classification problem from the point of view of the image content, where the disease manifests itself more as a variation in the leaf texture rather than a type of object that is present in the image, [3], [7], [9]. Such, different state-of-the-art CNN architectures for texture classification and recognition have been proposed [33]–[36]. However, this strategy is highly impacted by the lack of large texture datasets compared to the object classification problem datasets. Fine-tuning (retraining only some layers) on small texture datasets does not bring enough improvements in the classification accuracy [33]. In [33], the authors proposed the Texture CNN architecture which is based on AlexNet but uses an energy measure derived from the last convolutional layer. They arrived at the conclusion that the size of the dataset strongly influences the performance. The authors also observed that the fine-tuning performed on a network pre-trained on textured images achieves better results than by using a network pre-trained on a dataset that contains mostly objects. This happens probably because an image from an object-oriented dataset can contain multiple textures. In [36], the authors propose Bilinear CNN Models in which the fully connected layers are replaced with bilinear pooling models.

Our paper is structured as follows. Section II describes the proposed technique which involves using pre-trained CNN models on object-oriented datasets to extract textural features. Section III details the experimental configuration setups. We use publicly available textures and images of different real-world plant species affected by disease datasets for evaluation. Section IV details the obtained experimental results together with a comparison between other handcrafted and deep-learning methods and the proposed technique in terms of classification performance and time efficiency. Section V is dedicated to final conclusions and remarks.

II. THE PROPOSED METHOD

We are interested in the study of the performance of deep-learning pre-trained models in the classification of textured images even if the models were pre-trained on object categories. We show, therefore, how the chosen networks behave in a real task in which the textural characteristics are essential, namely in the classification of diseases that affect plant leaves. The underlying approach of the proposed method is to analyse which are the best pre-trained CNN models and their relevant layers for feature characterisation. We take advantage of the fact that there are large object datasets that allow for the pre-training of CNNs and keep the weights for the model and use this model in a new classification task.

The use of pre-trained models has several advantages. One of them is the fact that the feature extraction process is time-efficient because the images pass only once through the network. Secondly, relevant results can be obtained for small datasets for the classification task and no architecture handcrafting is required. This can be achieved because such models were trained on very large datasets, so there are many patterns and features already learnt that can be used to solve a different problem. For the significance of the results, the initial and the new task should be similar. Since we are interested in the classification of textures, even if the datasets on which popular CNN models were built are object-oriented, they are well-suited also for texture classification problems. This happens because of the hierarchical architecture of CNNs: while the early and mid-convolutional layers detect low-level features and texture structures, only the features computed from the last layers are more specific to the initial object classification task. We show in Fig. 2 the block scheme used to describe the considered texture classification system.

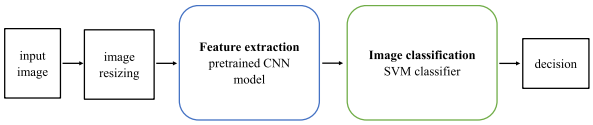

FIGURE 2. The block scheme of the considered classification system based on pre-trained CNNs.

We use a pre-trained CNN model from which a feature vector is obtained for each image of the dataset. The chosen supervised classifier is the Support Vector Machine with RBF kernel which is trained on 75% of the images from each class of the considered dataset and is evaluated on the rest (25%). In order to benefit from the advantages of the transfer learning concept and thus to keep the already learnt weights of the considered network, the classification layers at the end of the CNN network are removed because they are adapted to the number of classes on which the training of the CNN was performed, which is different to that of the current problem. Thus, pre-trained CNNs are used only for feature extraction in this work and the SVM is responsible for the classification. Although an artificial neural network consisting of a fully connected layer, a softmax layer, and an output layer could have been used for the classification part, it would not have surpassed the efficiency of SVM. According to [37], CNNs are very powerful as feature extractors due to their convolutional base, but less efficient for the classification operation since the classifier is in this case a linear one. On the other hand, SVM is better for the classification of more complex data [37] since by using the RBF kernel the initial feature space where data cannot be linearly separated is transformed into another higher dimensional space where the separation of data classes is possible. Using an SVM classifier on top of features extracted from CNNs instead of CNN classification layers provides better results in [38], [39].

The training and test sets are chosen randomly. For feature extraction, we considered several pre-trained models that are widely used in practice. They are presented in Table 1. Their default architecture is given in Appendix in Fig. 20-23. For these already existing models, the final classification layers are eliminated, and the features are extracted from several

TABLE 1. Considered pre-trained models.

Network	Number of convolutional and fully connected layers	Parameters (millions)	Size of the input image
ResNet18	18	11.7	224×224×3
AlexNet	8	61	227×227×3
Vgg16	16	138	224×224×3
ResNet50	50	25.6	224×224×3

different layers in order to observe which are the most relevant for a texture classification task.

All these models were pre-trained on the ImageNet object-oriented dataset which contains more than a million images of objects classified into 1000 classes. Each pre-trained model requires input images to be of a fixed size, as given in Table 1. So, if the analysed textured images have a different size, before using the pre-trained CNN models to extract the features, the input images are resized. The convolutional base performs convolutional operations by means of several filters. The weights are the filter values and they are determined by the number and size of filters. This means that the weights corresponding to the convolutional base network do not depend on the size of the input image. So, the convolution operation is not influenced by the input image size. Filter sizes remain the same if the input image size is changed. However, the size of the feature maps will be different and that is why the number of neurons for the fully connected layers is changed depending on the input image size. This means that retraining is necessary for this situation. Changing the architecture of the model would require changing the weights which is done by training and, in this case, the purpose behind the transfer learning concept would be lost. So, to be able to rely on this concept, the images are resized to match the size required by the considered CNN models. If the difference between the size of the initial images and that imposed by CNN models is not very large and the aspect ratio is kept the same (1:1), resizing the images does not bring artifacts that could negatively influence the performance.

We experiment with the extraction of features from several layers in the network. After feature extraction, the obtained feature vectors are fed into an SVM classifier whose parameters [40] are chosen through a grid search in order to obtain the best classification accuracy for each particular experiment and method.

III. EXPERIMENTAL SETUP

We validate the approach by two different experimental setups. Firstly, we investigate how transfer learning can be used in general for the classification of textures when the CNN models were pre-trained on large object datasets and what are, in practice, the relevant layers that can be considered from the hierarchical CNN to extract features from. Then, we use the results to provide an applied example of texture classification for the plant disease detection problem in precision agriculture.

A. TEXTURE DATABASE: OUTEX_TC_00013

For evaluating the proposed method, we used the Outex_TC_00013 dataset [41] which contains 68 categories of RGB textured images. There are 20 samples of size 128 × 128 pixels for each class, giving a total number of 1360 images. We show in Fig. 3 a sample for each image category. This dataset is challenging because the variability between different classes is rather small in some cases, such as the granite categories, the sandpaper ones, or the barleyrice classes. Therefore, the classification task can be difficult in such cases especially because the number of samples per class is limited.

B. PLANTVILLAGE DATASET

For validation of the method, we used the PlantVillage dataset [42] containing several plant species, some of them healthy and some affected by different diseases. In [43], the authors use three versions of this dataset: the original RGB images, the grayscale version, and the segmented RGB variant. In this paper, there is considered only the segmented RGB set. In our experiments, we only considered the segmented RGB images from [43] since the color information is relevant to this classification problem (as the change in leaf color can be a sign of a certain disease) and because the use of the segmented variant excludes any potential bias that might be caused by the presence of the background information. Images from this dataset were captured under different conditions, the plant leaves suffer different rotations and have different shapes. Moreover, there are some segmentation problems because the leaves are not always perfectly segmented from the background. We discarded from the initial dataset the images that were poorly segmented and could no longer be recognized. We show in Fig. 4 some examples of images that have segmentation problems, some of them being kept and some being discarded.

We performed two experiments: plant species identification and disease detection. For the plant species identification, we considered only the categories with healthy plant leaves. Fig. 5 shows three samples for each considered category for this experiment and Table 2 shows the 12 classes used in the plant species identification scenario. For the disease detection experiment, we considered several setups described in detail in Table 3. We also show in Fig. 6 some sample images for each class considered in each setup.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. OUTEX_TC_00013 RESULTS

In the first experiment, we considered extracting the features from the last layer located before the classification layers of the four pre-trained CNN models from Table 1: ResNet18, AlexNet, Vgg16, and ResNet50. For AlexNet and Vgg16, the last layer situated before the classification layers is fully connected (fc7 for both), whereas, for ResNet18 and ResNet50, the last layer is an average pooling layer (pool5 for ResNet18 and avg_pool for ResNet50). The pre-training of