Plant Classification using Convolutional Neural Networks

Hulya Yalcin, Salar Razavi
Visual Intelligence Laboratory
Electronics & Telecommunication Engineering Department
Istanbul Technical University
{hulyayalcin, razavi15}@itu.edu.tr

***Abstract* -- Application of the benefits of modern computing technology to improve the efficiency of agricultural fields is inevitable with growing concerns about increasing world population and limited food resources. Computing technology is crucial not only to industries related to food production but also to environmentalists and other related authorities. It is expected to increase the productivity, contribute to a better understanding of the relationship between environmental factors and healthy crops, reduce the labor costs for farmers and increase the operation speed and accuracy. Implementing machine learning methods such as deep neural networks on agricultural data has gained immense attention in recent years. One of the most important problems is automatic classification of plant species based on their types. Automatic plant type identification process could offer a great help for application of pesticides, fertilization and harvesting of different species on-time in order to improve the production processes of food and drug industries. In this paper, we propose a Convolutional Neural Network (CNN) architecture to classify the type of plants from the image sequences collected from smart agro-stations. First challenges introduced by illumination changes and deblurring are eliminated with some preprocessing steps. Following the preprocessing step, Convolutional Neural Network architecture is employed to extract the features of images. The construction of the CNN architecture and the depth of CNN are crucial points that should be emphasized since they affect the recognition capability of the architecture of neural networks. In order to evaluate the performance of the approach proposed in this paper, the results obtained through CNN model are compared with those obtained by employing SVM classifier with different kernels, as well as feature descriptors such as LBP and GIST. The performance of the approach is tested on dataset collected through a government supported project, TARBIL, for which over 1200 agro-stations are placed throughout Turkey. The experimental results on TARBIL dataset confirm that the proposed method is quite effective.**

Keywords— agriculture, plant classification, convolutional neural networks, deep learning, computer vision.

I. Introduction

Plants have an important role in human lives for many reasons such as growing concerns on lack of enough food resources in a world with exponentially increasing population and global climate change [1]. Automatic identification of plants based on spatiotemporal patterns extracted from their visual appearances draw attention of environmentalist as well. Traditional plant identification approaches are both expensive and time-consuming, since they require manual intervention of human experts [2]. Modern plant agriculture methods are immense since they have a great effect on the national economy and individual part of people's life [3].

Many countries across the world have been developing initiatives to build national agriculture monitoring network systems [4]. Similarly, an agricultural monitoring and information system has been established in Turkey through TARBIL project in 2012 [5]. As of 2016, over twelve hundred smart ground stations equipped with all sorts of sensors have been planted all over Turkey. The data collected through these stations are being accumulated on an online storage server and network infrastructure with fast fiber. Currently, a heavy manual human intervention is involved in identification and classification of plants through these systems. The accuracy of the conventional manual techniques depends on the observational skills and effort of the human observers where the uncertainty level and reliability of the yielding data cannot be guaranteed.

Recently, image analysis techniques have begun to emerge as an effort to automate the plant monitoring process [6]-[11]. Referring to the literature on plant classification, mostly approaches based on color features are used to develop a measure for plant identification. Color analysis usually relies on the distribution of colors in an image, but it is not a reliable feature since there are many conditions when the temporal consistency of this feature is violated. Illumination change, displacement of leaves with winds, camera jitter, zoom change, unexpected changes in camera parameters lead to inconsistent predictions of plant classification.

Despite numerous studies, plant classification based on digital images is still considered as a challenging problem. Some studies are based on analyzing individual plant leaves to identify and classify the plants [12]-[14]. Caglayan *et al.* have exploited color and shape features of the leaf images to classify plants [14]. Satti *et al.* have employed leaf images of the Flavia image dataset and applied k-Nearest Neighbor (k-NN) and Artificial Neural Networks (ANN) classifiers [15]. Gaber *et al.* applied Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA) to extract visual features from plants [13]. Various researchers have found the potential of large Convolutional Neural Networks to outperform traditional approaches for object recognition or detection that are based on ordinary color, texture, and shape features [16]-[19]. The CNN structures that have been used in these large-scale plant recognition tasks usually consist of a feature extractor followed by a classifier. All of the above mentioned methods deal with individual plant leaves. Obviously, acquiring plant leaf is not always an easy task because of the occlusions. Additionally, some plants may not even have recognizable leaf pieces. On the other hand, there have also been other studies related to employing computer vision methods to solve these problems. Landscape classification of aerial images using machine learning methods is an example of such applications of computer vision in agricultural computing [20]-[21].

Exploiting recent artificial intelligence methods like CNN to automatically classify plants is vital for improvement of the recognition accuracy. Classification is an active research area in pattern recognition which has been used in many applications that involves a decision mechanism that assigns a new observation to a set of classes, based on a training dataset. Traditional methods largely rely on manually crafting descriptors and characteristics. On contrary, in this paper, a Deep Learning (DL) method that can automatically extract features from two-dimensional plant images is proposed. The CNN models are an extension of deep learning of artificial networks. They consist of multi-hidden Layer Perceptrons (MLPs) which involve multiple convolution, pooling, ReLU, and fully-connected layers. Generally, feature maps of the prior layers are convolved with learnable weights at a convolutional layer and they are fed through the activation functions to form the output feature maps. Each output map combines convolutions with multiple input feature maps [12]. For a specific output map, the input maps will be convolved with distinct kernels and the convolutional layer shares its weights in a same output map. Convolutional layers are usually combined with pooling layers to reduce dimensions to increase speed and as a result further spatial variance is gradually built up. Although the pooling layer is desirable as it reduces the computational time, it may cause loss of some information. The pooling layer produces sub-sampled versions of the input maps. If there are *N* input maps, then there will exactly be *N* output maps, but the output maps will be smaller [12]. It sums over each distinct n-by-n block in the input matrix. For instance, if the down-sampling kernel is not over-lapping, the output matrix becomes n times smaller along both spatial dimensions. After processing through a combination of convolutional and pooling layers, the matrices are finally fed into fully-connected layers. In this layer, the parameters and weights are trained and learned by minimizing a loss function such as softmax and RBF [11]. Activation layers such as ReLU, sigmoid and tanh are used between input and fully-connected layers particularly for classification applications. It is often suggested vast amount of data is required for deep learning approaches to perform well. However, it is also important to use dropout, weight decay and data augmentation methods to deal with challenges like overfitting and increasing robustness.

In this paper, a Convolutional Neural Network (CNN) architecture is proposed to classify the type of plants from the image sequences collected from smart agro-stations. First challenges introduced by illumination changes and deblurring are eliminated with some preprocessing steps. Following the preprocessing step, Convolutional Neural Network architecture is employed to extract the features of images. In order to evaluate the performance of the approach proposed in this paper, the results obtained through CNN model are compared with those obtained by employing SVM classifier with different kernels, as well as feature descriptors such as LBP and GIST. The performance of the approach is tested on dataset collected through a government supported project, TARBIL, for which over 1200 agro-stations are placed throughout Turkey. The experimental results on TARBIL dataset confirm that the proposed method is quite effective. The rest of the paper is organized as follows: section II explains the proposed approach and experimental results are presented in section III.

II. PROPOSED APPROACH

We mainly use a pre-trained Convolutional Neural Network (CNN) model for classifying different plant species. In this section, we first explain the architecture of our CNN model and then report the details of how CNN model works in identifying different plant species. The overall framework and layers of our approach is shown in Table. 1. One of the most important challenges in dealing with high-dimensional data such as images is to connect the nodes of current layer to all the nodes in the previous layer. In our approach, instead of connecting each node in the current layer to the nodes in the previous layer, each node is only connected to a local region of the input volume. The spatial extent of this connectivity is called the receptive field of the connection. The convolution layer convolves the input layer with adjustable weight filters, namely kernels.

TABLE 1. THE ARCHITECTURE OF OUR CNN, EXPLICITLY SHOWING THE DELINEATION OF RESPONSIBILITIES OF THE GPU.

Layer	Sizes	Output †	Layer	Sizes	Output
Input	-	227×227×3	Conv 4	3×3×192	13×13×384
Conv 1	11×11×3	55×55×96	Relu 4	-	13×13×384
Relu 1	-	55×55×96	Conv 5	3×3×192	13×13×256
Norm 1	-	55×55×96	Relu 5	-	13×13×256
Pool 1	2×2×96	27×27 ×96	Pool 5	2×2×256	6×6×256
Conv 2	5×5×48	27×27×256	Fc 1 ††	6×6×4096	1×1×4096
Relu 2	-	27×27×256	Relu 6	-	1×1×4096
Norm 2	-	27×27×256	Fc 2	1×1×4096	1×1×4096
Pool 2	2×2×256	13×13×256	Relu 7	-	1×1×4096
Conv 3	3×3×384	13×13×384	Fc 3	1×1×4096	1×1×1000
Relu 3	-	13×13×384	Softmax	-	1×1×1000

† : width of the map×height of the map×# of color channels (or # of feature maps
†† : Fully-connected layer

Applying various types of weight filters, CNN attains shift invariant and scale invariant local features from the input. The pooling layer summarizes the output of the previous layer and achieves the invariance of translation. The purpose of pooling layer is to progressively decrease the spatial size of the data to reduce the amount of parameters and computation in the system, as well as to deal with overfitting problem. The pooling layer works independently on every depth slice of the input and resizes it spatially, using the max, min, medium or mean operations. In our approach, we employed max-pooling in each pooling layer. The local contrast normalization layer subtracts the means of neighborhood pixels from the target pixels and divides the subtracted pixels by the standard deviations of the local pixels. This layer allows the architecture to overcome the challenges introduced by various changes that occur between images captured under different conditions.

Conv1 is a feature extraction layer, and it takes its input from the input layer or sampling layer. Conv1 gets 96 two-dimensional feature maps for the size of 55 × 55 in the proposed approach. As a matter of fact, it is obtained by using an 11 × 11 convolution kernel with a stride of 4 and padding of 0. Each map of a convolution layer has the identical size as the convolution kernel. It has been observed that local features are not efficiently extracted when the convolution kernel is too small. On the other hand, the complexity of the extracted features may far exceed the ability of the convolution kernel when the convolution kernel is too large. Therefore, it is important to set the appropriate convolution kernel to increase the performance of CNN, as well as tuning the parameters of CNN. Pool1 layer is a sub-sampling layer and yields 96 feature maps with the size of 55 × 55. Sub-sampling layer is a layer that gets maximum value of the small region considered. In this paper, it sums up all non-overlapping sub-block X of 2 × 2 with a stride 2 pixels and padding 0. Then the sum is multiplied by weights ŵ and increased by an offset b. The sub-sampling operation is given by:

$$y = sigmoid(\hat{w}.sum(X_i) + b), \;\; for\, X_i \in X$$

Since the size of the feature map in Conv1 is 55 × 55, the final result for sub-sampling is 27 × 27. In Convolutional Neural Network, the general scaling factor is two. Reducing too fast corresponds to rough image feature extraction and loss of more features. Conv2 layer is another feature extraction layer. It has a similar functionality to that of Conv1, but there are some subtle differences in its filter size. Choosing a proper activation function is crucial since it significantly increases the performance of a CNN for certain applications. Rectified Linear Unit (ReLU) is one of the most commonly used non-saturated activation functions. The ReLU activation function is defined as follows:

$$y_i = max(0;\, z_i)$$

Here z_i is the input of i-th channel. ReLU is a piecewise linear function which prunes the negative part to zero and retains the positive part. The remaining layers of convolution and sub-sampling perform similar to those in previous layers. Only difference is that the extracted features become more abstract as the depth increases.

The CNN model learns and optimizes the filters in each layer through the back propagation mechanism. These trained and learned filters extract features that distinctively represent the input image. Therefore, instead of considering CNN as a black box, filter visualization is required to observe the transformation of the features backward, as well as to understand the internal operation and the characteristic of the layers and weights. In this work, models using stochastic gradient descent with a batch size of 128 examples, momentum of 0.9 (damping parameter), and weight decay of 0.0005 are trained as follows:

$$v_{i+1} = 0.9 \cdot v_i - 0.0005 \cdot \epsilon \cdot w_i - \epsilon \cdot \left(\partial L/\partial w|_{w_i}\right)$$

$$w_{i+1} = w_i + v_{i+1}$$

where ϵ is the learning rate and

$$\partial L/\partial w|_{w_i}$$

is the gradient of loss, L, with respect to the weight averaged over batch.

III. RESULTS AND DISCUSSION

The experimentation has been conducted on agricultural plants whose images are acquired by cameras mounted on agro-meteorological smart stations equipped with many sensors. As many countries across the world have been developing initiatives to build national agriculture monitoring network systems [4], similarly, an agricultural monitoring and information system has been established in Turkey through Turkish Agricultural Monitoring and Information Systems (TARBIL) project [5]. As of 2016, over twelve hundred smart ground stations equipped with all sorts of sensors have been planted all over Turkey. The data collected through these stations are being accumulated on an online storage server and network infrastructure with fast fiber. Collected images consist of close-up shots of certain areas of an agricultural field of interest. Every half an hour, images of the plants are taken and sent over to a server where thousands of images are accumulated in the database over the growth period of the various plants along with other information about the process. The plants that we experimented with were wheat, barley, lentil, cotton, pepper, corn, pomegranate, cherries, grapes, pistachio, tomato, apricot, sunflower, tangerine, beans and apple whose sample images are shown in Figures 1. It is not possible to get rid of unwanted effects of varying illumination conditions since many seasonal changes occur throughout the growing period of agricultural plants as can be seen in Figure 2 and Figure 3 where growth stages of pepper and corn are shown respectively.

We performed fine-tuning using a 16 classes from TARBIL dataset to build our Convolutional Neural Network (CNN). Thus, the final fully connected layer is set to have 16 neurons. As can be seen from Figures 1, 2 and 3, TARBIL dataset is quite challenging for classification. Not only plants from each class have different appearances throughout the growing stages, but also plants from different classes have similar color distributions. The provided training dataset consists of 4800 individual images that are organized into sixteen different classes, namely barley, sunflower, pepper, wheat, tomato, apple, bean, peanuts, apricot, cherries, tangerine, lentil, corn, pomegranate, cotton and grape. Each of these observations contain multiple images from the same plant, but the images belong to different growth stages of that plant.

The experiments were implemented in Matlab 2016a programming environment on a computer with Windows 7 Pro 64-bit operating system with following hardware specifications: Intel(R) Xeon (R) E5-1607, 16.00 GB RAM, NVIDIA Quadro K600. In order to evaluate the performance and efficiency of our deep-learning based approach, SVM based classifier. SVM classifier is experimented with RBF and polynomial kernels, as well as different kinds of features, namely four-connected LBP, GIST with 12 orientations per scale plus 4 blocks methods. In Table 2, the classification performance of our CNN based approach is compared with those of SVM classifier with different kernels and features. The accuracy rates on this dataset range from 69.81% to 97.47%. It should be recalled that CNN yields better results as the size of the data increases. As the numbers in Table 2 suggests, features learned from the CNN model outperforms state-of-the-art solutions that employ carefully chosen hand-crafted features. Even with different kernels and features, SVM classifier fails to achieve the performance of our CNN based approach. We also analyzed the drawbacks of our CNN model looking through misclassified patches. Most of the misclassified patches are from sunflower class with eighteen misclassified patches, followed by pepper class with nine misclassified patches, tomato class with eight misclassified patches, bean class with eight misclassified patches and apple class with six misclassified patches. Analyzing misclassified patches, the leading reason for misclassification turns out to be the appearance change of the plants due to the phenological changes and the illumination changes.

Fig 1. Sample images selected from 16 plant species used in our experimentation. Most of the plant species have similar color characteristics.

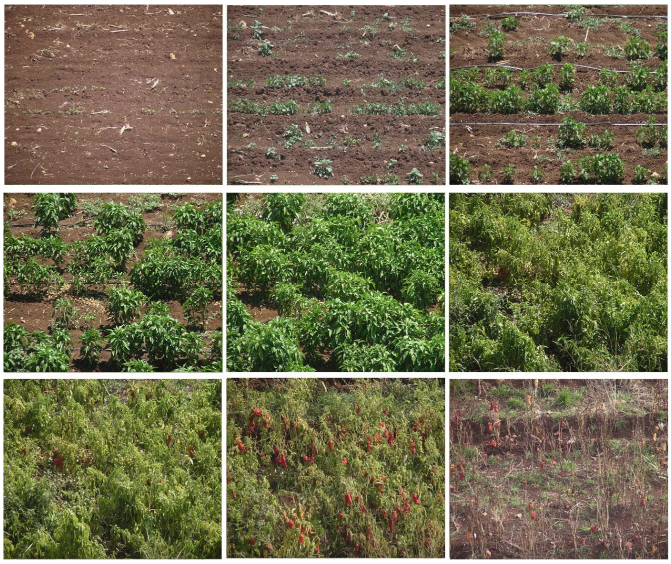

Fig 2. Different growth stages of pepper starting from plowing to cropping.

Fig 3. Different growth stages of corn starting from plowing to cropping.

TABLE 2. ACCURACY COMPARISON OF DIFFERENT SCHEMES.

Scheme	Feature	Accuracy (%)
SVM (RBF kernel)	LBP	74.92
	GIST	83.88
	Fusion	89.94
SVM (polynomial kernel)	LBP	69.81
	GIST	82.29
	Fusion	88.60
CNN	-	97.47

IV. CONCLUSION

In this paper, a convolution neural network based approach has been applied for the classification of a variety of plant species. Our CNN architecture can automatically classify images of sixteen kinds of plants. In order to evaluate the performance and efficiency of our deep-learning based approach, SVM based classifier using features such as LBP and GIST is also implemented. SVM classifier is experimented with RBF and polynomial kernels. The classification rate of these methods are compared to those of our CNN based approach. The algorithms were tested on the experimental data which were acquired under natural outdoor illumination. The data is provided by TARBIL Agro-informatics Research Center of ITU. Experimental results indicate that CNN based approach is significantly effective with an accuracy about 97.47% on 16 kinds of plants. Compared with other methods, experimental results suggest that the classification accuracy of CNN based approach outperforms other methods. Future work will consist of building different architectures, with a variety of activation functions, as well as experimenting pre-processing methods to enhance classification performance by improving the machine learning layer.

ACKNOWLEDGMENT

Data for this work is provided by Istanbul Technical University TARBIL Agro-informatics Research Center. This work is funded by the grants of Istanbul Technical University Scientific Research Fund (project # 34912), Turkish Ministry of Food, Agriculture and Livestock, and ITU TARBIL Agro-informatics Research Fund (project # 2012A020130).

REFERENCES

[1] Broich, M., Huete, A., "A spatially explicit land surface phenology data product for science, monitoring and natural resources management applications," Environmental Modelling & Software, 64: 191-204, 2015.

[2] Richardson, A., Keenan, T., Migliavacca, M., Ryu, Y., Sonnentag, O., Toomey, M., "Climate change, phenology, and phenological control of vegetation feedbacks to the climate system," Agricultural and Forest Meteorology, 169:156-173, 2013.

[3] Cleland, E.E.; Chuine, I.; Menzel, A.; Mooney, H.A.; Schwartz, M.D. Shifting plant phenology in response to global change. Trends Ecol. Evol., 22, 357–365, 2007.

[4] Srbinovska, M., Gavrovski, C., Dimcev, V., Krkoleva, A., Borozan, V., "Environmental parameters monitoring in precision agriculture using wireless sensor networks," Journal of Cleaner Production, 88:297-307, 2015.

[5] Bagis, S., Ustundag, B., Ozelkan, E., "An adaptive spatiotemporal agricultural cropland temperature prediction system based on ground and satellite measurements," International Conference on Agro-Geoinformatics, pages: 1-6, 2012.

[6] Xiao, X., Li, J., He, H., Fei, T., Zhou, Y., "Framework for phenology analyses from observation data of digital images and meteorological data," IEEE International Conference on Computer Science and Automation Engineering, pages: 373 - 377, 2012.

[7] Zhou, X., Luo, X., "Advances in Non-Destructive Measurement and 3D Visualization Methods for Plant Root Based on Machine Vision,"International Conference on Biomedical Engineering and Informatics, pp:1-5, 2009.

[8] H. Yalcin, "Phenology Monitoring of Agricultural Plants Using Texture Analysis," Signal Processing and Communications Applications Conference (SIU'16), Zonguldak, May 2016. (doi: 10.1109/SIU.2016. 7495926)

[9] S. Razavi, H. Yalcin, "Plant Classification using Group of Features," Signal Processing and Communications Applications Conference (SIU'16), Zonguldak, May 2016. (doi: 10.1109/SIU.2016.7496150)

[10] H. Yalcin, "Phenology Monitoring Of Agricultural Plants Using Texture Analysis," Fourth International Conference on Agro-Geoinformatics, Istanbul, 2015. (doi: 10.1109/Agro-Geoinformatics.2015.7248114)

[11] A. Vibhute and S. Bodhe, "Applications of image processing in agriculture: a survey," International Journal of Computer Applications, vol. 52, 2012.

[12] A. M. Coleman, "An adaptive Landscape classification procedure using geoinformatics and artificial neural networks," Vrije Universiteit, Amsterdam The Netherlands, 2008.

[13] T. Gaber, A. Tharwat, V. Snasel, and A. E. Hassanien, "Plant Identification: Two Dimensional-Based Vs. One Dimensional-Based Feature Extraction Methods," in 10th International Conference on Soft Computing Models in Industrial and Environmental Applications, Á. Herrero, J. Sedano, B. Baruque, H. Quintián, and E. Corchado, Eds., ed Cham: Springer International Publishing, 2015, pp. 375-385.

[14] A. Caglayan, O. Guclu, and A. B. Can, "A Plant Recognition Approach Using Shape and Color Features in Leaf Images," in Image Analysis and Processing – ICIAP 2013: 17th International Conference, Naples, Italy, September 9-13, 2013, Proceedings, Part II, A. Petrosino, Ed., ed Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 161-170.

[15] V. Satti, A. Satya, and S. Sharma, "An automatic leaf recognition system for plant identification using machine vision technology," International Journal of Engineering Science and Technology, vol. 5, p. 874, 2013.

[16] Z. Liu, L. Zhu, X.-P. Zhang, X. Zhou, L. Shang, Z.-K. Huang, et al., "Hybrid Deep Learning for Plant Leaves Classification," in Intelligent Computing Theories and Methodologies: 11th International Conference, ICIC 2015, Fuzhou, China, August 20-23, 2015, Proceedings, Part II, D.-S. Huang, K.-H. Jo, and A. Hussain, Eds., ed Cham: Springer International Publishing, 2015, pp. 115-123.

[17] T. J. Jassmann, R. Tashakkori, and R. M. Parry, "Leaf classification utilizing a convolutional neural network," in SoutheastCon 2015, 2015, pp. 1-3.

[18] N. Sünderhauf, C. McCool, B. Upcroft, and T. Perez, "Fine-Grained Plant Classification Using Convolutional Neural Networks for Feature Extraction," in CLEF (Working Notes), 2014, pp. 756-762.

[19] C. Zhang, P. Zhou, C. Li, and L. Liu, "A Convolutional Neural Network for Leaves Recognition Using Data Augmentation," in Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), 2015 IEEE International Conference on, 2015, pp. 2143-2150.

[20] T. Oberth, J. Cock, M. S. Andersson, R. N. Naranjo, D. Casta, et al., "Acquisition of low altitude digital imagery for local monitoring and management of genetic resources," Comput. Electron. Agric., vol. 58, pp. 60-77, 2007.

[21] D. Dey, L. Mummert, and R. Sukthankar, "Classification of plant structures from uncalibrated image sequences," in Applications of Computer Vision (WACV), 2012 IEEE Workshop on, 2012, pp. 329-336.

[22] U. Weiss and P. Biber, "Plant detection and mapping for agricultural robots using a 3D LIDAR sensor," Robotics and autonomous systems, vol. 59, pp. 265-273, 2011.