Machine Learning-based inference system to detect the phenological stage of a citrus crop for helping deficit irrigation techniques to be automatically applied.

Forcén M.[1], Pavón, N.[1], Pérez D.[1], Berríos P.[1], Pérez-Pastor A.[1], López J.A.[1]

1. Universidad Politécnica de Cartagena, Spain

Motivation

- Enabling smart irrigation systems to be more efficient in terms of saving water and other resources by applying intelligent techniques for visual analysis.
- Laying the ground work of the use of machine learning algorithms to extract features and characteristics from image data of crops.

Objectives

In this work, deep learning techniques are applied for phenological stage prediction of a citrus crop using data from a camera.
Additionally, several parameters of the network will be fine-adjusted with the aim of analysing the performance impact in the final prediction system.

Materials and methods

Several pictures of citrus crop trees at different stages are taken (see Fig. 1). A standard RGB camera is used (digital camera with sensor IMX363 Exmor RS).
These images are classified manually, using the standard criteria of phenology stages, as stated in [1].

Figure 1. One of the images part of the dataset

This learning system has been implemented by using Tensorflow [2], the Deep Learning framework made by Google.
Images are preprocessed by the Tensorflow API (Application Programming Interface) in order to be suitable to use in a Deep Learning architecture.
Each one is divided into 8 pieces, ignoring the original image borders. The pieces are then scaled to a size of 256x256 pixels.
Finally, the preprocessed image data is introduced in the following neural network whose architecture is defined in Figure 2.

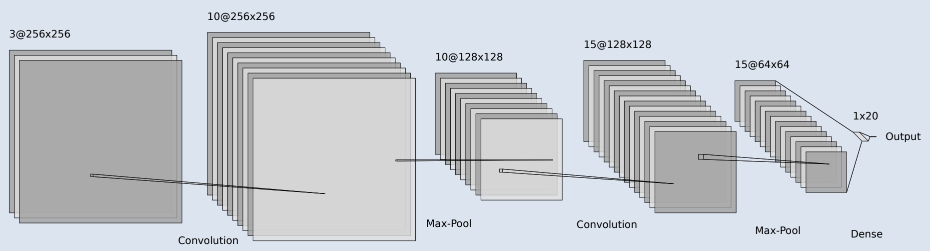

Figure 2. Main deep learning network architecture

Since the aim of the network is to obtain a continuous value instead of a discrete one, the problem to solve is a regression.
The optimization algorithm used to train the neural network is the *Adam* ([3] algorithm.
Another important choice is the metric that is used to minimize the error of the network. In this case, the MSE[1] is used.
Then, some parameters of the network, such as the numbers of neurons of the last dense layer or the number of filters (along with their sizes), will be empirically adjust for testing their impact in accuracy and efficiency of the system.
Finally, a collection of images included in a validation set (which are different from the training set), will be used to calculate the phenological stage in order to test the prediction accuracy.

[1] Mean Square Error

Training and test

First, parameters involved in the training stage will be properly defined, by using empirical techniques for adjusting them.
Given the available data (304 pictures of trees), two datasets will be generated, one for training, with 200 pictures, and another one for testing, which includes the rest of the images.
Each training process consists of 10 epochs. This will ensure that the network extracts all the necessary features to conduct the regression procedure accurately.
During the first stage, the number of hidden neurons is a parameter to be tuned. Values from 20 to 100 in steps of 10 neurons have been tested. Tuning results are shown in Figure 3.

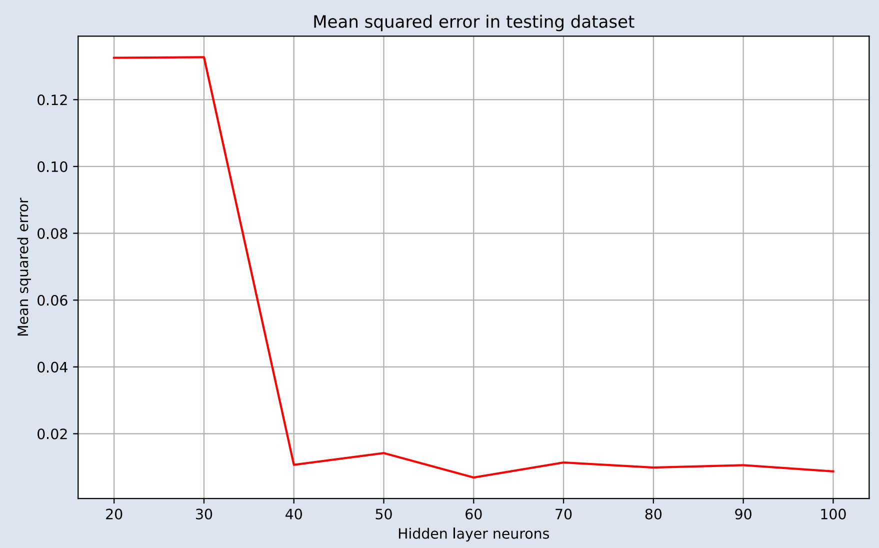

Figure 3. MSE in testing dataset vs hidden layer neurons

While the final values are different between runs of the training processs, the main guidelines remain constant. It is remarkable that adding more neurons is useless at some point as there is no major impact on the prediction accuracy, but the required calculation get more complex.
The second experiment made in this work has been focused on analysing the importance of the number of convolutional layers and their parameters. For this purpose, the same training process is executed with one layer and two layers. The previous results are taken into account, keeping the last dense layer with 40 neurons. For each variation of the architecture, 10 training process were run. After each training, the test dataset was used to obtain an indicator of the accuracy of the network. The values are then shown in Table 1.

Convolutional layers	1	1	2
Convolutional filters	10	10	10→ 15
Filter size	4	8	4→ 5
Mean of MSE	0.1111	0.0396	0.0137
Variance of MSE	0.0054	0.0025	0.0000

Table 1. MSE statistics vs neural network architecture.

It is noticeable that the one layer architecture results in a significantly less accurate network. Furthermore, it is possible that the network converges in a local minimum, decreasing the performance in those runs.

Conclusion

This work demonstrates the feasibility of using machine learning and deep learning techniques in order to extract plant features from images with several degrees of success.
The architectures developed in this work will be a suitable starting point for other projects that rely on these techniques.

Finally, several neural networks architectures has been analysed and compared in order to define the minimum structural requirements to get accurate results and to prove the possible effects caused by an insufficient architecture.

References

[1]A. Bain et al., «Plasticity and diversity of the phenology of dioecious Ficus species in Taiwan», Acta Oecologica, vol. 57, pp. 124-134, may 2014 doi: 10.1016/j.actao.2013.10.004.
[2]Martín Abadi et al., TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015.
[3]D. Kingma y J. Ba, «Adam: A Method for Stochastic Optimization», International Conference on Learning Representations, dic. 2014.

Acknowledgements

This work was funded by the Ministerio de Investigación e Innovación del Gobierno de España, through the Water4Ever European Project (reference PCIN-2017-091, WaterWorks2015 ERA-NET), and the fundings of the predoctoral contract for Manuel Forcén, thanks to Fundación Séneca (reference 20767/FPI/18)