Precise Weed and Maize classification through Convolutional Neuronal Networks

Andrea Córdova-Cruzatty *, Mauricio Daniel Barreno Barreno † and José Misael Jácome Barrionuevo‡
Department of Energy and Mechanics, University of the Army ESPE
Latacunga - Quijano y Ordoñez and Hermanas Paéz
Email: *accordova@espe.edu.ec, †mdbarreno@espe.edu.ec, ‡jmjacome1@espe.edu.ec

***Abstract*—Deep Learning has played an important role in modern times, being widely used in artificial vision applications and specifically in pattern recognition. The versatility of deep learning has positioned it as a fit tool used in many fields of application, among which is precision agriculture. This paper presents a study of a convolutional neural network (CNN) with two classes applied for the classification between maize plants and weeds in real time, focused on maize crops in their initial development stages. The algorithm has two stages, segmentation of images and classification. Segmentation is intended to separate the target plant from the original image, while classification is meant to identify what images belong to the two classes by means of a convolutional neural network previously trained with a dataset generated through the segmentation stage. The performance of the network was analyzed and compared with different network architectures. The network architecture that presented the best training results was chosen and afterwards its number of filters were reduced in order to analyze its behavior. Finally, the resulting network was tested on different processors to compare classification speed of each and determine which one could be used for an autonomous system in a real-time application.**

I. Introduction

During the last centuries, significant progress has taken place in science and technology developments. Significant milestones in diverse areas such as communications, numerical computer control and the miniaturization of components have benefited industrial sectors, providing new means to approach problems in a different manner.

The availability of leading edge technologies has considerably increased, thus becoming more accessible to people. This effect of globalization has allowed many countries to have access to state-of-the-art technological products, promoting its application in different areas to generate alternative and more efficient solutions to conventional problems.

Technological advances have caused industry transformations; manufacturing, food processing, and information industries among others, have gone through dramatic changes with the new industrial revolution. However there are fields that have not been significantly influenced by technological development in Ecuador, as it is the case of agroindustry. Agriculture in Ecuador has undergone some changes since colonial times; while it is true that there are efficient agricultural practices, the lack of technological resources make it impossible for the country to exploit its true potential as a clean and efficient agricultural producer. Also, the excessive use of pesticides have a negative effect on the production, soil, and water quality, which leads the new research efforts towards finding alternative solutions to weed control issues.

Nowadays, one of agriculture's challenges is the development of precision agriculture techniques focused on weed and crop segmentation. There are studies that show the impact of weed in corn crops [1]; its yield is affected by 5000 kg/ha. Currently, growing development of artificial vision and machine learning algorithms has allowed researchers to propose solutions for weed segmentation in different types of crops. These algorithms are intended to be used in different weed control systems, such as autonomous weed-removing robots.

One of the first approximations for crop detection algorithms was developed in 1996 [2]; this algorithm could segment crops from weeds by means of infrared (IR) Images. The images were processed by a hysteresis umbral and the Min Neighbouring method to identify the rows of a crop. In recent years the implementation of machine learning has opened new possibilities in weed and crop differentiation process. Recently, the use of Harris Corner detection techniques in CNNs was implemented with the use of Feature Detection, a machine-learning algorithm that uses DBSCAN (Density-based spartial clustering of apllications with noise) [3]. This algorithm demonstrated an effectiveness of 98% in the identification of rice weeds. Araguez *et. al* [4] performed their segmentation through the analysis of the green histogram and performing the crop and weed segmentation with unspecified classifiers in their document (revisar si esta bien 2 ultimas oraciones?).

Hong and Lei [5] developed an approximation by using an optimal method for detection in various luminosity settings, they achived this by the use an artificial neural network (ANN) for weed and maize classification with a precision of 92.5%. Also, another algorithm that uses OTSU and Watershed binarization methods for image segmentation has been developed [6]. While classification in this algorithm was done through an area analysis to perform a thresholding and the method is computationally effective when the weed distribution does not resemble the size to the crop plants, its error increases when crop density is higher compared to weeds. Romeo *et. al* [7] proposed to use a fuzzy clustering approach to correctly segment green within the crop.

Weed and crop segmentation is not limited to color images. In another application [8], the use of a a multispectral camera permitted to obtain RGB and NIR images for the segmentation and classification purposes. The images used a light CNN for the first process and a deep CNN for classification, its accuracy was as high as 98% in the identification of weeds. One of the most difficult things in crop identification is the generation of datasets due to weed presence in the training images, but these can be correctly generated by the use of a CNN [9], achieving an acceptable accuracy for the generation of datasets (verdad?).

II. Matherials and Methods

The developed model for weed and maize identification algorithm using convolutional neural networks is described below.

A. Hardware

For the training and image-capturing stages, the following elements were used: a Raspberry Pi 3 with a V2.1 Pi camera for image acquisition. The system was configured to obtain a video at a resolution of 1280x720 pixels in order to get more detail in the segmented images. Due to the popularity of GPU computing for its advantages in the development of models of Convolutional Neural Networks, such as a better processing speed, it was necessary to consider hardware compatible with parallel computing devices such as the Graphic Cards manufactured by NVIDIA. These cards have the advantage of a superior processing speed compared to a CPU in order to optimize image-processing time and deep learning Model's training process, thus the selected device was a computer with a Nvidia graphics card GTX950M. Also, it was necessary to consider the use of a Core i7 2.7 Ghz CPU of 8 cores and the CPU of the Raspberry Pi 3, with an ARM Cortex-A53, 1.2GHz, and 4 cores, in order to test their performance for a future application in an independent mobile system that could be implemented in an embedded system. The hardware chosen will be tested to classify images, in each classification a measurement of the processing time will be taken to later on analyze if the hardware can process information in real time.

B. Software

The use of powerful libraries of image processing and deep learning development was required due to..(ver espanol).. . The use of open software was required in order to provide the system with a certain level of flexibility to extend its use for future prototypes. Thus, OpenCV library was chosen due to its computational efficiency and focus on real-time applications [10]; it was used specificaly for sample images acquisition and segmentation. Also, Caffe was considered because it is intended to aid in the training and developing of general-purpose convolutional neural networks and other deep lerning models [11] with full support of CUDA GPU computation library and a generous database of resources for learning of applications such as simple regression, speech, robotics, large-scale visual classification, etc.

Also, the use of Linux OS in the development stages was important due to its light operating system and free OS which is also fully compatible with OpenCV. The computer used in the project worked with Ubuntu Linux 16.04 due to its official support to CUDA, Caffe and OpenCV. Likewise, Pixel OS was used in Raspberry Pi for acquiring images and testing the CNN classification performance.

C. Dataset Description

Also, it was necessary to obtain images of the most common weed plants found in maize crops in Pillaro, in order to allow the network to discriminate between the two defined classes in the dataset: maize and weed.

Then, through a first-stage digital image processing procedure, the images were normalized in its green channel. This was done in order to improve green color detection based on the elimination of light and shadow in the images. Later, the green color was extracted from the images to obtain a grayscale equivalent; this was done through the following equation described by Wang, Meng, Luo and Mei [12]. Following, an OTSU Thresholding was applied to the grayscale images in order to obtain a binary mask.

$$S = 2 * G - R - B \tag{1}$$

Afterwards, the images were segmented in order to differentiate them from the soil and other non-plant elements. Original images were masked with the binary mask to provide the dataset with color features of maize and weed plants. Once the segmented images were obtained, the files that had a resolution lower than 64x64 pixels were removed because its size was smaller than the size of the network input layer. Besides, the processing of these small images was not convenient since its total area was much smaller than the maize sample images. Following, the images were saved in a png format. The entire process described before is illustrated in the next figure.

It should be noted that samples were manually labeled into two classes; maize and weed, making sure that the images are not repeated. In the following image, final results of the processed images that conform the dataset are shown.

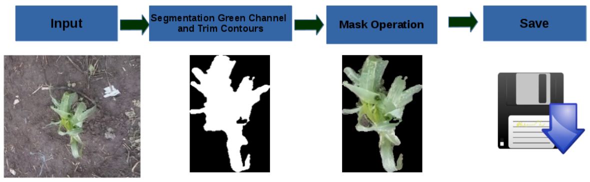

Fig. 1: Preprocessing to obtain samples for the dataset

Fig. 2: Segmented and Masked images of dataset Above: Weed, Below: Maize

Once image processing was completed, a set of 2835 maize images and 880 weed images was obtained. Lee *et. al* [13] said that once CNN has reached an acceptable accuracy, the dataset can be extended by means of geometric transformations. In addition, geometric transformations [14] allow overfitting reduction and improve precision in the training process. For this reason, initially obtained images were rotated every 30 degrees, obtaining an increment of the dataset by 12 times, thus increasing the chances of effectively recognizing plants that are in any orientation.

For the Validation Phase, a fifth part of the total image set of each class was chosen, resulting in the following chart with the distribution of the dataset.

Phase	Train	Validation	Test	Total
Maize	25695	8325	200	34220
Weed	8560	2000	200	10760

TABLE I: Distribution of the Dataset of each class

D. *Convolutional Neuronal Network*

After the dataset was completed, it was neccesary to classify maize and weed images. Convolutional Neural Networks (CNN) can be used as a highly-accurate method for image classification, these models are complex but efficient with a high rate of discrimination and have proven to provide good results in image classification, object detection, and fine-grained classification [15]. CNN are widely used in precision agriculture [8] for the correct identification of plants.

One of the characteristics of CNN is that these can have multiple architectures, each architecture can reach different results depending on the application. For the present document four CNN architectures have been considered. Two arquitectures were chosen from the Caffe Zoo Model(a free resource provided by Caffe Developers), LeNet and AlexNet. Also, Potena and Nardi used two additional types of architectures, sNET and cNET, which sucessfully proved an aceptable performance in plants and crop identification processes[8]. With the four Nets chosen, each one was trained with the same solver of type AdaDelta and with the same dataset. The training results are the following.

Parameters	LeNet	AlexNet	cNET	sNET
Input size of images	32x32	64x64	64x64	64x64
Layers numbers	9	11	8	4
Number of parameters	652500	20166688	6421568	135872
Iterations	3000	3000	3000	3000
Accuracy(%)	86.48	93.86	96.4	80.4
Loss(%)	32.80	15.32	13.72	15.32

TABLE II: Comparison of the 4 types of CNN

Based on the performance parameters shown on the table II, the network that achieved the highest rate of accuracy and lowest rate of loss was selected; clearly the cNET presented a superior performance due to its near-to-human classification precision.

Although results given by this type of network are excellent, one of the main purposes of this paper was the implementation of this system in real time. The number of parameters to be computed by the selected network were numerous, so network modifications in order to improve processing time had to be made. To achieve this, it was necessary to reduce the number of filters of cNET from 64 to 16.

Parameters	cNET 16 filters	cNET 64 filters
Number of parameters	1651376	6421568
Iterations	9000	9000
Accuracy(%)	97.26	96.40
Loss(%)	8.39	14.48

TABLE III: Comparison between cNET of 16 and 64 filters

This filter reduction permitted to decrease the number of parameters to 1651376 (this represents only 25% of the original network parameters) without compromising the network precision; it still reached a 97.2% of accuracy.

Thus, the complete network model was conformed by the following components: an input layer where an RGB image of 64x64 pixels is received, followed by a convolution layer of 16 filters, a kernel of 5 and a stride of 1 pixel (this layer only obtains the main characteristics of the image), then a pooling layer is applied, which has the function to find the maximum values of the previous layer with a value of kernel size and stride of 2. Then again a second layer of convolution that has 16 filters is applied, the size of kernel is 5 and the stride is 1, followed by a new layer of pooling with a kernel size and stride of 2, so the resulting image had a size of 16x16x16. Finally, the network is completed with three consecutive fully-connected layers; the first with 384 neurons, the next with 192 and the last with 2 neurons. The type of layers fully-connected contain a total of 478 neurons, where the last two neurons that are the output represent the two classes to identify, the result model is shown in the next graph.

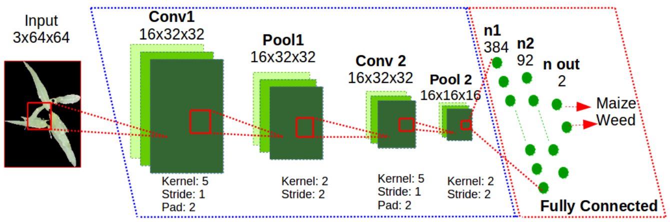

Fig. 3: The schematic of cNET

In the training graph shown below, the final net accuracy result can be appreciated; 9000 iterations were necessary to reduce the loss value to its minimum and the architecture shows that it can reach an acceptable accuracy within few iterations. Note that by reducing the number of filters the network gained accuracy because the dataset does not have too many features to process, so the network obtained a better solution with a fewer number of parameters.

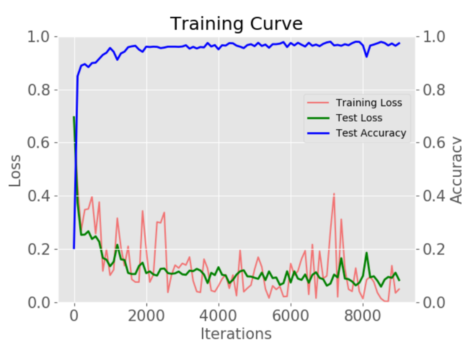

Fig. 4: Graph of the Training Process

III. Test

For this stage, another dataset was considered with 202 images of maize plants in stage V3-V7 and 202 images of weeds, obtained in different maize crops than the ones used for gathering the training dataset images. Ten measurements of the mean classification time of the test dataset were obtained for each one of the different hardwares previously selected to run performance tests. Classification tests were started with Caffe without any software modifications. The obtained results were compared between cNET arquitecture with 64 and 16 filters, these are shown in the next Table IV.

GPU			
cNET	**Accuracy Weed**	**Accuracy Maize**	**Average time of classification/image**
64 Filters	95.05%	74.75%	2.34 ms
16 Filters	94.55%	87.13%	951 μs
CPU			
cNET	**Accuracy Weed**	**Accuracy Maize**	**Average time of classification/image**
64 Filters	95.05%	74.75%	51.6 ms
16 Filters	94.55%	87.13%	17.23 ms
CPU Raspberry Pi 3			
cNET	**Accuracy Weed**	**Accuracy Maize**	**Average time of classification/image**
64 Filters	95.05 %	74.75%	615 ms
16 Filters	94.55%	87.13%	161 ms

TABLE IV: Test with the three types of Hardware

As the table shows, the best results in classification time were achieved by the GPU processor, thus it is the best option to classify a large amount of images. However, this project is thought to be implemented in an autonomous system so the Raspberry Pi was selected for the classification performance analysis due to its portability. The resulting classification time for the cNET of 64 filters was 615 miliseconds, however this time is not optimum for an approximation of real time classification.

The results shown previously are not definitive because Caffe has ways to optimize image dataset classification. Multithreading was applied to this research project with the help of the mathematical software OpenBLAS. Also, batching was used to classify each class dataset in a single forward pass. For each hardware, Caffe was recompiled with OpenBLAS instead of Atlas. In the program, all images were entered in a vector to be classified in an only batch per each class (ACLARAR IDEA). Also, OpenBLAS was configured to use the 4 cores of each hardware to obtain even results, the obtained results are the following.

Hardware	**Accuracy Weed**	**Accuracy Maize**	**Average time of classification/image**
GPU	89.11%	92.08%	1.58 ms
CPU	89.11%	92.08%	10.92 ms
CPU(Raspberry Pi)	89.11%	92.08%	150.8 ms

TABLE V: Test of cNET 16 filters with optimized software

The results show that for the GPU, classification time has decreased by 66 %, mainly because the OpenBLAS Library makes an extensive use of the CPU threads, causing that not all parameters will be computed by GPU as in the previous test. On the other hand, when the CPU is used the results are better, because the operations are executed using each of the four cores. In the case of the CPU, notice a 22 % and for 6.3

% for the Raspberry Pi.(REVISAR)

To compare all these results taking into account that an image captured by the camera at a height of 1.5 m can focus an approximate 6 plants of maize, and according to the amount of weed, an average of 12 plants, Giving a total of about 18 plants. This analyzes how many frames per second would be appropriate for the acquisition of images.

Parameter	**GPU(Without Threading)**	**CPU with Threading**	**RaspBerry Pi 3 with Threading**
Time(s)	0.0171	0.196	7.714
FPS	58.47	5.08	0.36

TABLE VI: Test of a complete image and the frames per second

IV. Conclusion

In this paper we have studied the application of Convolutional Neuronal Networks for Classification between Maize and Weed Plants, for that we have used some CNN Architectures, the Network with best results in precision was cNET, with a 97.23 % of accuracy of training, using a dataset of both classes of 44580 segmented images.

Also, by using different methods to classify the images in our experiments, the results showed that the best time of classification was achieved by using a Nvidia Graphic Cards, this is due the Parallel Architecture that this GPU has and also it has fully support for the computation platform CUDA, it demostrated to be 18 times faster than a normal CPU and 170 times faster than a Raspberry Pi 3.

The study demostrated that is better to perform some geometrical tranformations to the Dataset images due it benefits to improve the classification of plants because the plants ever have different orientations in the crop field.

-The use of Multithreading and Batching through the library OpenBLAS is very beneficial in the case of using CPU, increasing the speed of processing in the percentage higher than 20 %, with the Rasberry pi, a minimum improvement was obtained, instead it was pejudicioal with the use of GPU, because it causes that the speed is reduced considerably.

Acknowledgment

The authors would like to thank to the ”Granja Agroecológica y Demostrativa” of Píllaro City and owners of corn crops of the city which allowed us to take samples of plants, to conform the dataset of this project.

This work has been supported by the University of the Army ”ESPE” Latacunga.

References

[1] R. SUÁREZ and J. P. Y. J. VALLADARES, “Distintos sistemas de escarda en maíz forrajero,” *Producciones agroganaderas: Gestión eficiente y conservación del Medio Natural. Actas de la XLV RC de la SEEP. Gijón*, 2005.

[2] R. Brivot and J. Marchant, “Segmentation of plants and weeds for a precision crop protection robot using infrared images,” *IEE Proceedings-Vision, Image and Signal Processing*, vol. 143, no. 2, pp. 118–124, 1996.

[3] B. Cheng and E. T. Matson, “A feature-based machine learning agent for automatic rice and weed discrimination.” in *ICAISC (1)*, 2015, pp. 517–527.

[4] S. Equipo Técnico, “Proceedings of vii seae congress spanish society for organic farming,” 2006.

[5] H. Y. Jeon, L. F. Tian, and H. Zhu, “Robust crop and weed segmentation under uncontrolled outdoor illumination,” *Sensors*, vol. 11, no. 6, pp. 6270–6283, 2011.

[6] S. H. Hlaing and A. S. Khaing, “Weed and crop segmentation and classification using area thresholding,” *IJRET*, vol. 3, pp. 375–382, 2014.

[7] J. Romeo, G. Pajares, M. Montalvo, J. Guerrero, M. Guijarro, and A. Ribeiro, “Crop row detection in maize fields inspired on the human visual perception,” *The Scientific World Journal*, vol. 2012, 2012.

[8] C. Potena, D. Nardi, and A. Pretto, “Fast and accurate crop and weed identification with summarized train sets for precision agriculture,” in *International Conference on Intelligent Autonomous Systems*. Springer, 2016, pp. 105–121.

[9] M. Di Cicco, C. Potena, G. Grisetti, and A. Pretto, “Automatic model based dataset generation for fast and accurate crop and weeds detection,” *arXiv preprint arXiv:1612.03019*, 2016.

[10] G. Bradski, “The opencv library.” *Dr. Dobb’s Journal: Software Tools for the Professional Programmer*, vol. 25, no. 11, pp. 120–123, 2000.

[11] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, “Caffe: Convolutional architecture for fast feature embedding,” *arXiv preprint arXiv:1408.5093*, 2014.

[12] P. Wang, Z. Meng, C. Luo, and H. Mei, “Path recognition for agricultural robot vision navigation under weed environment,” in *7th International Conference on Computer and Computing Technologies in Agriculture (CCTA)*, no. Part I. Springer, 2013, pp. 242–248.

[13] S. H. Lee, C. S. Chan, P. Wilkin, and P. Remagnino, “Deep-plant: Plant identification with convolutional neural networks,” in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 452–456.

[14] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, “Deep neural networks based recognition of plant diseases by leaf image classification,” *Computational intelligence and neuroscience*, vol. 2016, 2016.

[15] A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, “Cnn features off-the-shelf: an astounding baseline for recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2014, pp. 806–813.