Precise Weed and Maize Classification through Convolutional Neuronal Networks

Córdova-Cruzatty Andrea *, Barreno Barreno Mauricio Daniel † and Jácome Barrionuevo José Misael ‡
Departamento de Energía y Mecánica, Universidad de las Fuerzas Armadas ESPE
Sangolquí - Ecuador
Email: *accordova@espe.edu.ec, †mdbarreno@espe.edu.ec, ‡jmjacome1@espe.edu.ec

***Abstract*—Deep Learning is playing an important role in big data processing for more accurate modeling of common productive processes. It is 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 the development of an algorithm capable of image segmentation 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 defined classes. It applies a convolutional neural network (CNN) to discriminate maize plants from weeds in real time, at early crop development stages. It was applied to maize crop because it is a common staple crop in the Ecuadorian Highlands. The convolutional neural network has been trained with a dataset generated in the segmentation stage. The performance of the network was analyzed with LeNET, AlexNet, cNET and sNET network architectures. The network architecture that presented the best training results was cNET based on its performance in terms of accuracy and processing time. The minimum working filter number for this network architecture was 16. The best performing algorithms and processors have a significant potential for autonomous weed and crop classification systems in a real-time application.**

Keywords*—Machine Learning, Convolutional Neural Networks, GPU Computation, Caffe, Segmentation of plants, Classification of maize.

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 and productive sectors [1], 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 [2], 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, agricultural industries among others have gone through dramatic changes with the new industrial revolution such as the Precision agriculture(PA) which has opened a new field for researching [3]. 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 and applications make it impossible for the country to exploit its true potential as a clean and efficient agricultural producer, for the next years it will be planned to introduce some changes concerning to the innovation to improve production [4]. 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 aspects.

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 [5]; yield can be affected up to 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 [6]; this algorithm could segment crops from weeds by obtaining a set 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 aplications with noise) [7]. This algorithm demonstrated an effectiveness of 98% in the identification of weeds in rice crops. Araguez *et. al* [8] implemented a segmentation algorithm through the analysis of the green histogram of the image of plants and performed the crop and weed segmentation with unspecified classifiers.

Hong and Lei [9] developed an approximation by using an optimal method for detection in various luminosity settings, they achived this by the use of 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 [10]. While classification in this algorithm was done through an area analysis to perform a thresholding and the method is computationally efficient when the weed distribution does not resemble the size of the crop plants, its error increases when crop density is higher compared to weeds. Romeo *et. al* [11] 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 [12], the use of 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 [13], achieving an acceptable accuracy for the generation of datasets.

II. Matherials and Methods

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

A. Hardware

A Raspberry Pi 3 with a V2.1 Pi camera for image acquisition was used for the training and image-capturing stages. 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, 1.2 GHz, 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 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 because the images had to be processed with artificial vision algorithms and classified with a selected convolutional neural network architecture. 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 [14]; it was used specifically 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 [15] 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 the software chosen. The computer used in the project worked with Ubuntu Linux 16.04 due to its official support to CUDA, Caffe and OpenCV. Likewise, PIXEL Distribution derived from Debian Linux was used in Raspberry Pi for acquiring images and testing the CNN classification performance.

C. Dataset Description

Due to the lack of local weed and maize plants datasets, it was necessary to search for those classes of plants in maize crops. The initial development stage of maize was selected because actions taken by the farmer during this stage are easier to execute and these will impact the future crop yield. Fieldwork at a nearby town called Píllaro was arranged with the purpose of obtaining image samples. Píllaro is a city located in the Tungurahua Province, in the center of the Ecuadorian Highlands region, it is a recognized andean-crop producer of species such as maize, potatoes and a wide variety of fruits. Due to its year-long production season, many maize crops at its initial development stage were found in this area.

Samples were obtained from images captured in different maize fields. Crops where weed and maize plants could be visually discriminated were chosen in order to easily segment the images and provide reliable samples to the network, which allow the model to correctly learn all features of the real plant. Maize plants at a V3 to V7 development stage (3-7 leafs) [16] were used to take samples. To record the samples, a camera was centered over the target plant, insuring that the captured image showed all its features. This was achieved by monitoring the obtained samples through an external display. Examples of the images captured are shown in Fig. 1.

Fig. 1. Crops chosen to obtain samples, it is easy to discrimante weed from maize

Also, it was necessary to obtain images of the most common weed plants found in maize crops in Píllaro, in order to allow the network to discriminate between maize and weeds.

Then, through a first-stage digital image processing, 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 (1) described by Wang, Meng, Luo and Mei [17]. 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 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 Fig. 2.

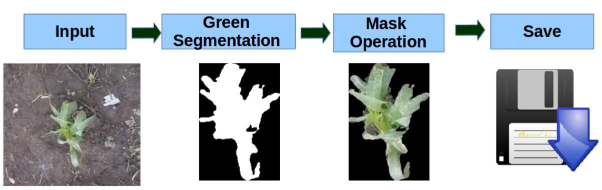

Fig. 2. Preprocessing to obtain samples for the dataset

It should be noted that samples were manually labeled into two classes; maize and weed, making sure that the images are not repeated. In Fig. 3, final results of the processed images that conform the dataset are shown.

Once image processing was completed, a set of 2835 maize images and 880 weed images was obtained. Lee *et. al* [18] said that once CNN has reached an acceptable accuracy, the dataset can be extended by applying geometric transformations. In addition, geometric transformations [19] allow overfitting reduction and improve precision in the training process. For this reason, the 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.

Fig. 3. Segmented and masked images of dataset Above: Maize, Below: Weed

For the validation phase, randomly the fifth part of the total image set of each class was chosen, resulting in the dataset distribution shown in Table I.

TABLE I
DATASET DISTRIBUTION OF EACH CLASS

Phase	Train	Validation	Test	Total
Maize	25695	8325	202	34222
Weed	8560	2000	202	10762

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 [20]. CNN are widely used in precision agriculture [12] for the correct identification of plants.

One of the characteristics of CNN is these 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 [12] used two additional types of architectures, sNET and cNET, which sucessfully proved an aceptable performance in plants and crop identification processes. With the four Nets chosen, each one was trained with the same solver of type Adam and with the same dataset. The training results are shown in Table II.

Based on the performance parameters shown in Table II, the network that achieved the highest rate of accuracy and lowest rate of loss was selected; clearly the cNET presented a

TABLE II
COMPARISON OF THE 4 TYPES OF CNN

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

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 is the implementation of a real time system. The number of parameters to be computed by the selected network were numerous, so it was necessary to make network modifications in order to improve its processing speed. To achieve this, it was necessary to reduce the number of filters of cNET from 64 to 16. This final number of filters used in the network was selected because this is the minimum number of filters that maintain an acceptable training accuracy. The results of the experiment are shown in Table III.

TABLE III
COMPARISON BETWEEN CNET OF 16 AND 64 FILTERS

Parameters	cNET 16 filters	cNET 64 filters
Number of parameters	1651376	6421568
Iterations	9000	9000
Accuracy(%)	97.26	96.40
Loss(%)	8.39	13.72

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 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 fully-connected layers contain a total of 478 neurons, where the last two neurons are the output representing the two classes to identify. The resulting model is shown in Fig. 4.

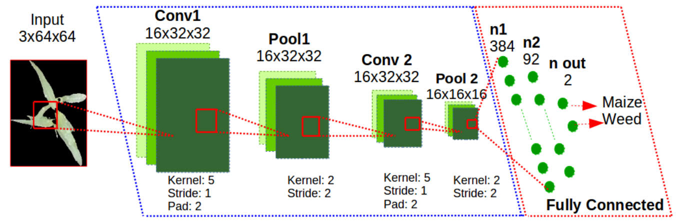

Fig. 4. The schematic of cNET

In the training graph shown in Fig. 5, 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.

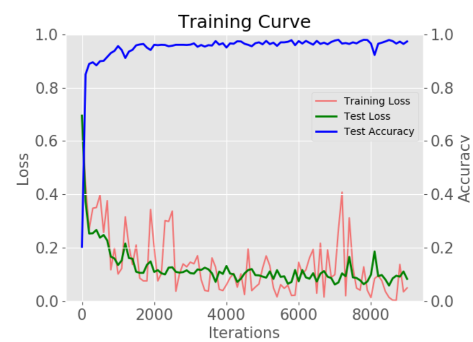

Fig. 5. Graph of the Training Process

III. TEST

For this stage, another dataset was considered with 202 images of maize plants in stage V3-V7 and 202 images of weeds. 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 in vanilla version. The obtained results were compared between cNET arquitecture with 64 and 16 filters, these are shown in Table IV.

As the Table IV 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 16 filters was 161 miliseconds, however this time is not optimum for an approximation of real time

TABLE IV
TEST WITH THE THREE TYPES OF HARDWARE

GPU			
cNET	**Accuracy Weed**	**Accuracy Maize**	**Average time of classification/image**
64 Filters	95.05%	74.75%	2.34 ms
16 Filters	92.08%	89.11%	951 μs
CPU			
cNET	**Accuracy Weed**	**Accuracy Maize**	**Average time of classification/image**
64 Filters	95.05%	74.75%	51.6 ms
16 Filters	92.08%	89.11%	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	92.08%	89.11%	161 ms

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 with cNET of 16 filters, all images were entered in a vector to be classified in an only batch per each class. Also, OpenBLAS was configured to use the 4 cores of each CPU hardware to obtain even results among these two devices. Since GPU has thousands of cores, this test could not be configured under the same restriction for this device. The obtained results are the shown in Table V.

TABLE V
TEST OF cNET 16 FILTERS WITH OPTIMIZED SOFTWARE

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

The results show that for the GPU, classification time has increased by 66%, mainly because the OpenBLAS Library makes an extensive use of the CPU threads, causing some parameters to be computed in the CPU instead of the GPU, thus making the process slower. For the CPU the results are better because the operations are executed using each of the four cores, for this case a decrease in time of 22% was obtained. Finally, the result for the Raspberry Pi CPU was a time decrease of 6.3%.

Lastly, an image captured by the camera at a height of 1.5 m can focus an approximate 6 plants of maize. Taking into account the number of weeds, an average of 12 plants can show up in the image. The total number of plants per image would round 18 plants considering maize and weeds as well. This aids to determine how many frames per second would be appropriate for the acquisition of images.

For the estimation of the frames per second(FPS) for each device, its optimal configuration was considered. The GPU classification time during the test without multithreading was considered for this device. For the CPU and the Raspberry Pi CPU, the results for classification time in the multithreading tests were considered. The devices that showed a realistic number of frames per second are the GPU and the CPU, while the Raspberry Pi CPU did not obtain an achievable rate like itś shown in Table VI.

TABLE VI
TEST OF COMPLETE IMAGE CLASSIFICATION IN FPS

Parameter	**GPU(Without Threading)**	**CPU (Threading)**	**Raspberry Pi (Threading)**
Time(s)	0.0171	0.196	2.714
FPS	58.47	5.08	0.36

IV. CONCLUSION

In order to have an adequate image processing that could be implemented to work in real time, the number of network filter layers was reduced from 64 to 16, reducing the number of parameters to be computed during the classification. A classification time reduced by 2.5 times compared the original Net of 64 filters was obtained. This change had a small impact in the training accuracy of the net, its value increased 1 % in comparison to the original net.

This paper presented an analysis of a Convolutional Neuronal Network application for classification between maize plants and weeds. With this purpose, several CNN Architectures were tested, the network that presented the best results was cNET of 16 filters, with a 97.23 % of training accuracy, using a dataset of both classes of 44580 segmented images.

Also, by using different methods to classify images, the results showed that the best classification time was achieved by using a Nvidia Graphics Card. This was due to the parallel architecture that this GPU has. Also, this Graphics Card is fully supported for the computational platform CUDA, it demonstrated to be 18 times faster than a normal CPU and 170 times faster than a Raspberry Pi 3.

The study demonstrated that performing geometrical transformations to the dataset images improves plant classification accuracy since the plants will have different orientations in the crop field.

The use of Multithreading and Batching through the library OpenBLAS is beneficial when using a CPU, increasing the processing speed higher than 20 %. In contrast, a minimum improvement was obtained with the Raspberry Pi. On the contrary, the use of a GPU was prejudicial because it reduces the processing speed considerably.

For implementing the algorithm in an embedded system, it is not recommended to work with a processor such as the Raspberry Pi 3 at the moment. This embedded system has a very slow processing speed because the frames per second used to analyze the images are significantly low, so a card that has GPU should be chosen, without the need for Multithreading and Batching. Even though there are promising advances in external Deep Lerning devices for embedded systems such as the Intel Movidius Neural Compute Stick, these systems are not yet available.

Acknowledgment

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

This work has been supported by the Universidad de las Fuerzas Armadas ESPE.

References

[1] J. P. Davim, *Microfabrication and Precision Engineering: Research and Development*. Woodhead Publishing, 2017.

[2] G. M. Grossman and E. Helpman, "Globalization and growth," *The American Economic Review*, vol. 105, no. 5, pp. 100–104, 2015.

[3] J. Ortiz-Cañavate and P. Elorza, *LAS MÁQUINAS AGRÍCOLAS Y SU APLICACIÓN*. Mundi-Prensa, 2012. [Online]. Available: https://books.google.com.ec/books?id=OF2VMQEACAAJ

[4] F. Houtart, "El desafío de la agricultura campesina para el ecuador," *La restauración conservadora del correísmo*, p. 167, 2014.

[5] 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.

[6] 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.

[7] 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.

[8] S. Equipo Técnico, "Proceedings of vii seae congress spanish society for organic farming," 2006.

[9] 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.

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

[11] 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.

[12] 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.

[13] 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.

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

[15] 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.

[16] A. Fassio, "Maíz, aspectos sobre fenología, alberto fassio, ana inés carriquiry, cecilia tojo, ricardo romero," *Serie técnica. 101.*, 1998.

[17] 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.

[18] 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.

[19] 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.

[20] 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.