ISSN 2007-9737

Corn/Weed Plants Detection Under Authentic Fields based on Patching Segmentation and Classification Networks

Francisco Garibaldi-Márquez[1,2], Gerardo Flores[1], Luis M. Valentín-Coronado*[,1,3]

[1] Centro de Investigaciones en Óptica A. C., Guanajuato, Mexico

[2] Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias–Campo Experimental Pabellón, Pabellón de Arteaga, Mexico

[3] Consejo Nacional de Humanidades, Ciencias y Tecnologías, Mexico

{franciscogm, gflores, luismvc}@cio.mx, garibaldi.francisco@inifap.gob.mx

Abstract. Effective weed control in crop fields at an early stage is a crucial aspect of modern agriculture. Nonetheless, detecting and identifying these plants in environments with unpredictable conditions remain a challenging task for the agricultural industry. Thus, a two-stage deep learning-based methodology to effectively address the issue is proposed in this work. In the first stage, multi-plant image segmentation is performed, whereas regions of interest (ROIs) are classified in the second stage. In the segmentation stage, a Deep learning model, specifically a UNet-like architecture, has been used to segment the plants within an image following two approaches: resizing the image or dividing the image into patches. In the classification stage, four architectures, including ResNet101, VGG16, Xception, and MobileNetV2, have been implemented to classify different types of plants, including corn and weed plants. A large image dataset was used for training the models. After resizing the images, the segmentation network achieved a Dice Similarity Coefficient (DSC) of around 84% and a mean Intersection over Union (mIoU) of around 74%. On the other hand, when the images were divided into patches, the segmentation network achieved a mean DSC of 87.48% and a mIoU of 78.17%. Regarding the classification, the best performance was achieved by the Xception network with a 97.43% Accuracy. Then, According to the results, the proposed approach is a beneficial alternative for farmers as it offers a method for detecting crops and weeds under natural field conditions.

Keywords. Deep learning, weed detection, segmentation and classification, corn field variabilities.

1 Introduction

Corn holds great gastronomical and economic significance for many countries across the globe. In Mexico, for instance, the sown area has kept steadily in the last decade (2010 – 2020), with an average sown surface of 8 million hectares annually.

Nonetheless, the demand for this cereal increased by 136% in the same period [3], which has been compensated with importations. In this sense, factors such as land tenure, weather change, and crop management could avoid the self-sufficiency of this cereal for the country.

Among management practices, weeds elimination is one of the most important tasks in agriculture because these unwanted herbs compete with crop plants for nutrients, sunlight,

ISSN 2007-9737

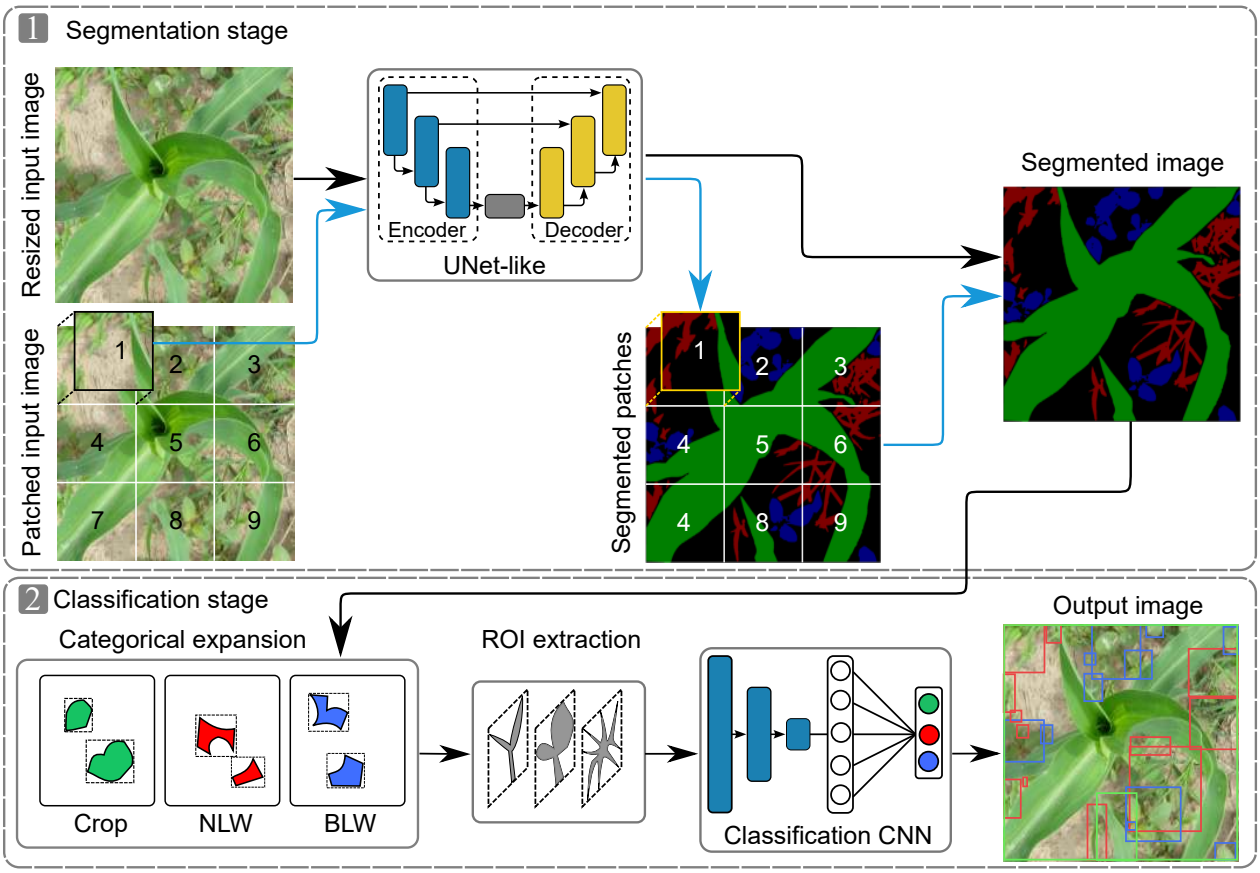

Fig. 1. Propose a method for detecting crop and weed plants in authentic corn fields, utilizing segmentation and classification networks.The resulting image output comprises of green, red, and blue boxes, each representing the Crop, NLW, and BLW classes, respectively

and water [19] and could lead to 90% of kernel yield reduction if not controlled in time [13]. The most commonly employed control strategy to eradicate weeds from cornfields is through the application of herbicides.

However, the excessive use of herbicides has resulted in environmental pollution [9]. This is predominantly due to the uniform application of significant volumes of these chemicals throughout the entire field, even in regions where weeds are absent [11]. Consequently, to address the environmental impact of herbicides while sustaining crop yield, researchers have developed a sophisticated technique termed site-specific weed management (SSWM). This method involves the targeted application of chemicals exclusively in areas where weeds are present, thereby minimizing environmental pollution.

Operating systems that can effectively distribute adequate herbicides on individual weed plants or patches of them in the fields is plausible [14]. Nevertheless, detecting (localizing and classifying) these plants in natural crop environments has been reported to be a demanding and intricate task [6]. This challenge is primarily attributed to diverse parameters, such as the intensity of sunlight, the density of plants, foliage occlusions, and the variety of plant species.

The implementation of Convolutional Neural Networks (CNNs) for identifying crop and weed plants has gained significant traction in recent times. YOLO [11] and Faster-RCNN [10] are

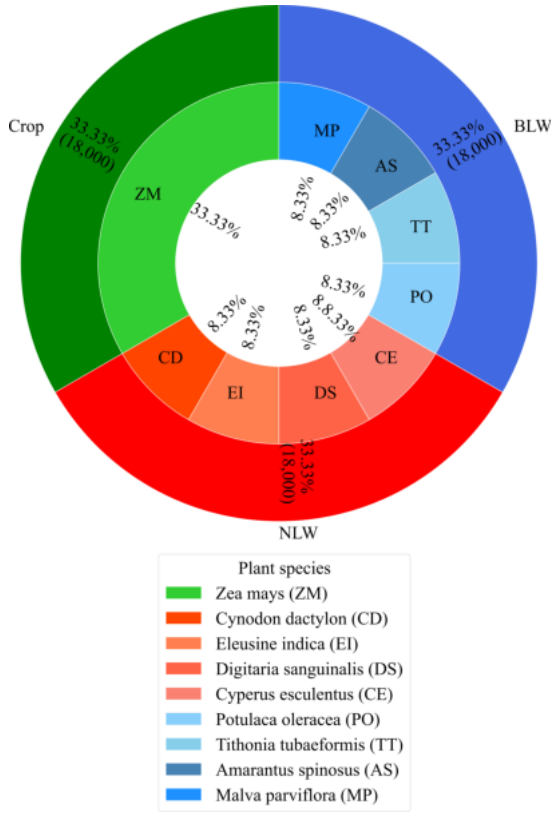

Fig. 2. Visualization of the experimental dataset showcasing plant species grouped into distinct classes, with corresponding labels meticulously traced per individual plant species

among the popular architectures employed for this purpose. However, their efficacy is limited when detecting plants in densely populated fields.

A promising alternative is the technique of semantic segmentation, which separates the plants from the background, although it requires additional algorithms for classification.

While established architectures exist for the semantic segmentation of objects within images, there has been limited research on weed segmentation in corn fields, mainly due to the unavailability of a large and diverse corn/weed dataset.

Here, we use deep learning models to segment and classify corn and weed plants under authentic environments and high plant density.

1.1 Related Works

The segmentation of plants in natural conditions poses a significant challenge due to the complexity of the variables involved. These variables include the plant species, density, foliage occlusion, morphological changes across growth stages, soil appearance, and sunlight intensities.

The presence of these variables makes it challenging to extract and classify the unique features of plants. Few works in the literature have been conducted on the segmentation of weeds in corn crops. However, Fawakherji et al. [5] recently proposed a method for segmenting a multispectral dataset.

The images were captured using an unmanned aerial vehicle (UAV) within a natural cornfield environment and classified into soil and green plants. A VGG-UNet model was then trained using four sub-dataset images derived from Red, NIR, synthetic images from the Normalized Difference Vegetation Index (NDVI), and RED+NIR+NDVI.

Results showed a mean accuracy of 73%, 85%, 92%, and 88%, respectively. It is worth noting that multispectral channels offer better segmentation performance compared to the visible spectrum [2, 12]. However, the associated cost of infrared sensors would present a challenge for autonomous weed control systems.

Visible spectrum cameras have been utilized in discriminating between corn and weeds in real fields. For instance, in the work of Quan et al. [16], the segregation of weeds under complex cornfield environments was explored using the BlendMask network. An extensive dataset of 5,700 images was formed, which included two broadleaf weeds and one narrowleaf weed.

Results indicated that a ResNet101 backbone yielded a higher mIoU of 60.7% compared to 50.2% with ResNet50. More recently, Picon et al. [15] employed the PSPNet network in segmenting a corn/weed dataset in natural fields, resulting in a Dice Similarity Coefficient (DSC) of 25.32%.

This dataset consisted of corn, narrowleaf weed (three species), and broadleaf weed (three species). However, the authors acknowledged that

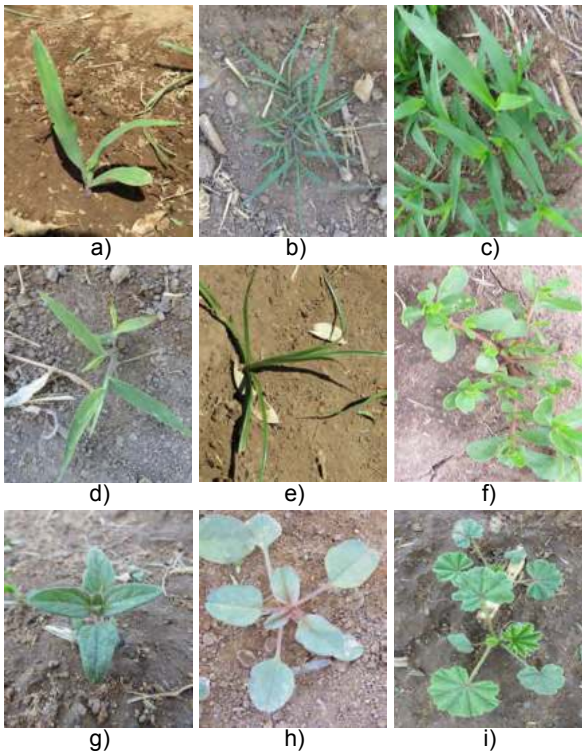

Fig. 3. Sample of the plant species of the experimental dataset. a) Zea mays, b) Cynodon dactylon, c) Eleusine indica, d) Digitaria sanguinalis, e) Cyperus esculentus, f) Portulaca oleracea, g) Tithonia tubeoformis, h) Amarantus spinosus and i) Malva parviflora

the narrowleaf class was not correctly classified due to its visual similarity to the crop class.

In this work, we present a large dataset of corn and weed images that were captured in authentic natural corn fields. This dataset includes four monocotyledon plant species and four dicotyledon plant species as weeds, as well as corn plants as the crop.

To detect the Crop, narrowleaf weeds (NLW), and broadleaf weeds (BLW), we propose a deep learning-based approach. Each weed class, NLW and BLW, groups the four plant species of weeds, respectively. The proposed approach performs well despite the challenging conditions presented in the acquired images.

The rest of the document is structured as follows: Section 2 contains the dataset description as well as the implementation details of the segmentation approaches. Section 3 presents the primary results of the experiments, and Section 4 provides the conclusions of the work.

2 Materials and Methods

According to the results obtained from our previous work [7], the UNet-like model [17], whose encoder layer was the network ResNet101, performs the segmentation of plants adequately. However, it has been observed that the model often misclassifies the pixels of the isolated Regions of Interest (ROIs).

This evidenced the necessity of developing a vision system with the ability to detect corn plants, narrowleaf weeds, and broadleaf weeds under authentic corn fields, giving the excessive field variabilities. This gap is covered by proposing a detection method based on deep learning segmentation and classification networks, as shown in Figure 1.

The algorithm comprises two main stages: a segmentation stage and a subsequent classification stage. In the segmentation stage, an image with multiple plants is segmented using a UNet-like architecture. The segmentation process has been carried out under two approaches.

In the first approach, the input images are segmented in a simple step by simply resizing them, whereas in the second approach, the input images are divided first into patches to avoid the loss of significant features, and then each patch is segmented. Subsequently, in the classification stage, the pixels belonging to each class (Crop, NLW, or BLW), from the segmented image are first separated into single-class images, and then each image is transformed into binary masks for the easy extraction of the ROIs under scenarios of high density of plants.

These ROIs are extracted using the well-known connected component analysis (CCA) [8]. Then, in the final stage, an image is obtained within the detected plants that have been detected. To perform this task, the networks ResNet101, VGG16, Xception, and MobileNetV2 have been implemented and evaluated. The implementation details of the segmentation

Table 1. Metrics adopted for evaluating the UNet-Like model

Name	Acronym	Definition
Dice Similarity Coefficient	DSC	$\frac{2\ \mathrm{TP}}{2\ \mathrm{TP}+\mathrm{FP}+\mathrm{FN}}$
Intersection over Union	IoU	$\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}}$
Mean Intersection over Union	mIoU	$\frac{1}{N}\sum_{j=1}^{N}\mathrm{IoU}_j$

and classification networks are covered in Sections 2.2.1 and 2.2.2, respectively.

2.1 Dataset Description and Image Pre-Processing

The dataset consisted of 12,000 visible spectrum images captured from five corn fields in Aguascalientes, Mexico. Three corn fields were established during the spring-summer agricultural cycle of the year 2020, and two additional corn fields in the same cycle of the year 2021.

The dataset images have varying dimensions, including 4,608 $\times$ 3,456 pixels, 2,460 $\times$ 1,080 pixels and 1,600 $\times$ 720 pixels. During the process of capturing images, the camera was positioned at a distance between 0.4 m and 1.5 m above the soil surface. Consequently, a significant number of images were captured from a top-down perspective, while a limited number had a side view.

Furthermore, it is noteworthy that most top-down view images were captured from a distance greater than 1 m to avert dust accumulation on the camera lens, which can be caused by agricultural tractors traveling through crop fields.

It is, therefore, recommended that during the tentative instrumentation, the camera should be positioned at a height of more than 1 m from the ground to avoid such issues. The dataset contains various factors that introduce variability.

The plants' variability is determined by the number of species, instances in a single image, and occlusion and foliage overlap. Changes in zoom and side views also affect the scale and perspective of the plants.

Furthermore, the dataset includes plants in different growth stages, starting from two true leaves to seven true leaves, captured every five days. Soil status is another parameter that affects the dataset, including humidity conditions, organic matter content, and changes in its appearance, such as color and texture. The images were captured in different sunlight intensities, including morning, noon, and evening, as well as on sunny and cloudy days.

After integrating the dataset, meticulous manual annotation of each image at a pixel level was conducted. The aim of this process was to precisely quantify not only the crop species (Zea mays L.), but also eight different weed plant species: four narrow-leaf weeds (NLW) and four broadleaf weeds (BLW). Figure 2 summarized the plant species and the labels traced per each of them. Noticed that they have been grouped into the classes Crop, NLW, and BLW. Furthermore, Figure 3 shows a sample of the plant species of the dataset.

To develop an effective detection strategy, a Convolutional Neural Network (CNN) was trained using a sub-dataset comprising of individual-plant images that were extracted from the original experimental dataset's multi-plant images. This approach ensured that the dataset used for training the classification networks was well-balanced, with 18,000 images per class.

2.2 Training of the Architectures

The detection approach proposed involves two stages, as previously mentioned. The first stage employs a UNet-like network for the segmentation process. The second stage involves implementing and evaluating ResNet101, VGG16, Xception, and MobileNetV2 networks for the classification process. The CNN architectures were trained on a desktop computer that boasted a Core i7 processor, 32 GB of RAM, and an NVIDIA GeForce RTX 3070Ti GPU with 8GB of memory.

The implementation was carried out in Python 3.8, utilizing the Keras framework with Tensorflow 2.5.0 as the backend.

Table 2. Metrics adopted for evaluating the classification models

Name	Definition
Accuracy	$\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}$
Precision	$\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}$
Recall	$\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$
F_1-score	$2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$

2.2.1 UNet-Like

First of all, to train the UNet-like proposed model, a fine-tuning of the hyper-parameter was performed. Therefore, a few training steps were implemented before figuring out the better configuration of UNet-like network. In a first trial, the encoder and decoder blocks were trained, and their weights were randomly initialized.

Then, in a second trial, a transfer learning strategy was implemented to the network, i.e., the weights of the convolutional layers of ResNet101 (encoder) were imported from that when it was trained in the ImageNet dataset [4], and then they were frozen. The learning rate, the optimizer, and the number of epochs also were changed.

In the segmentation approach where all input images are segmented in a single step, it was only necessary to resize the image and train the network. On the other hand, an image padding pre-processed was implemented in the approach in which the input images were divided into patches.

Thus, the original size of input images remains unchanged, and pixels of value 0 were added on two sides of them to obtain fixed-size patches. The loss function always was the dice loss, since it is very strict for segmentation tasks because it penalizes those predominant pixels of certain classes.

The computation of dice loss is as follows:

$$L_{\text{Dice}} = 1 - \frac{2\,y\,y^* + 1}{y + y^* + 1}, \tag{1}$$

where y and y^* refer to the ground truth and the predicted model value, respectively.

2.2.2 Classification Networks

In all the cases, the convolutional layers of the classification networks were the original from the architectures, but the Fully Connected (FC) layers were proposed. Then, we have established the parameters and hyper-parameters of these architectures, following a similar approach to that of the segmentation network. Firstly, the weights of the convolutional and FC layers were initialized randomly and trained. Secondly, the convolutional layers were initialized with weights obtained from the ImageNet dataset and subsequently retrained with our own dataset.

In this step, only the FC layers were trained. Furthermore, we have changed the FC layers from two to three. Thus, the neural network architecture employed in our study consisted of variable numbers of neurons, ranging from 512 to 4,096, with increments of 512 for the first and second layers. The ReLu activation function was used for the first two layers, while the output, which was the third layer, comprised three neurons with a softmax activation function.

This choice of activation function was motivated by the three specific classes of our dataset, namely Crop, NLW, and BLW. To optimize the neural network's performance, we employed a fine-tuning process that involved varying the optimizer, learning rate, loss function, and number of epochs. This approach allowed us to achieve superior results and ensure the accuracy of our model.

2.3 Evaluation Metrics

The proposed approach was evaluated in two stages. The initial stage involved the assessment of the segmentation process, followed by an evaluation of the classification stage.

The chosen evaluation metrics for the UNet-like segmentation network are Dice Similarity Coefficient (DSC), Intersection over Union (IoU), and mean Intersection over Union (mIoU). These metrics have been selected to assess the

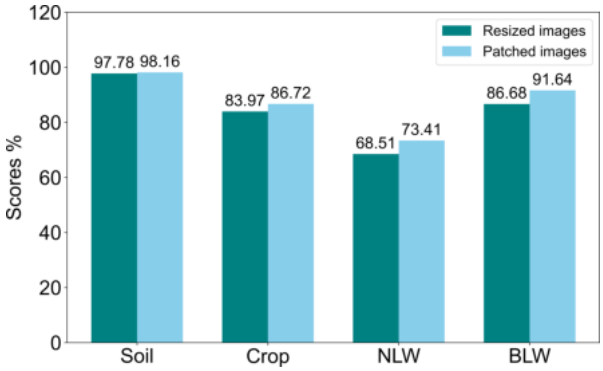

Fig. 4. DSC achieved by the UNet-like model when the input images were resized and divided into patches

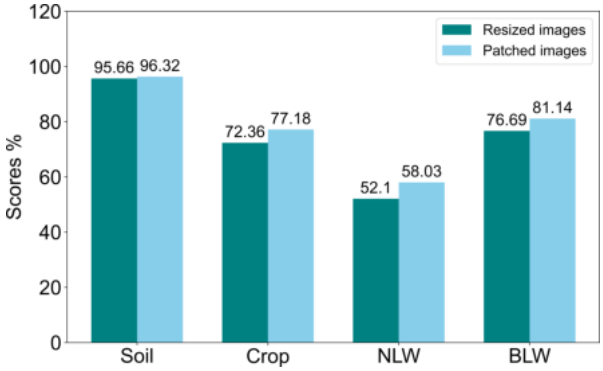

Fig. 5. IoU reached by the UNet-like model when the input images were resized and divided into patches

network's performance and provide an accurate representation of its effectiveness.

DSC pixel-wise compares the similarity between the ground truth and the predicted mask, reflecting their size and localization agreement as perceptual quality [1]. IoU is employed to calculate the percentage of overlap and align concerning the desired outcome.

The metrics utilized to evaluate the networks' performance are presented in Table 1. Noticed that the $mIoU$ is computed considering the total number of classes (N) of the dataset.

The performance assessment of our classification models was conducted using the established metrics of Accuracy, Precision, Recall, and F_1-score. Table 2 offers an insightful overview of these metrics. In Table 1 and 2, the TP (true positive), TN (true negative), FP (false positive), and FN (false negative) values are directly estimated from the confusion matrix.

3 Results and Discussion

This section provides an overview of the results obtained from the segmentation network UNet, as well as the classification networks' performance. Furthermore, we will undertake a comprehensive analysis of the achievements of each task. A set of representative images showcasing the accurate detection of crop and weed plants is also presented to understand the system output better.

3.1 Performance of the Unet-Like Model

The segmentation stage has been carried out under two approaches. The first one consists of segmenting the resized input images, whereas in the second approach, the input images are divided first into patches, and then each patch is segmented.

In either case, the best results were obtained when the transfer learning technique was implemented to train the UNet-like model. Regardless of the approach, the network input image size was 512 $\times$ 512. In addition, and according to the experimentation, the Adam optimizer with a learning rate of 0.0001 was observed to fit better into our dataset. The number of epochs used to train the model was 100.

The performance of the DSC metric of the trained UNet-like model, when images were resized and divided into patches is depicted in Figure 4. It is observed that the four classes of the dataset were better segmented by the UNet-like model when the images were divided into patches since the DSC of the four classes is superior under this scenario. Specifically, the BLW class was found to be better segmented by the network, followed by the Corn class, and finally, the NLW class, when focusing solely on the plant classes.

A narrow analysis indicates that the classes Crop, NLW, and BLW were 2.75%, 4.90% and 4.96% better segmented respectively when images were divided into patches, in contrast when they were resized and segmented in a step.

ISSN 2007-9737

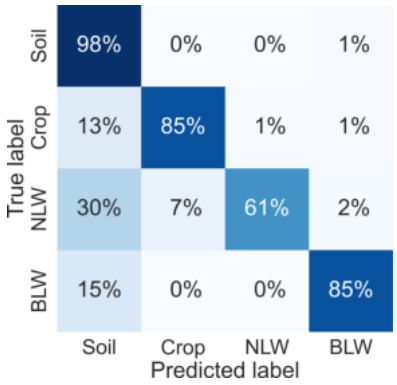

Fig. 6. Confusion matrix obtained when the images were solely resized for segmenting

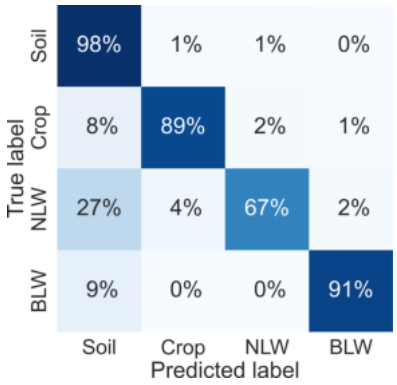

Fig. 7. Confusion matrix obtained when the images were divided in patches for segmenting

The same behavior of the UNet-like model is observed for the metric IoU under the two segmentation scenarios, as Figure 5 shows. That is, the UNet-like model performs better for all the classes when images are divided into patches. The IoU reaffirms that the BLW was the best-segmented class, then the class Crop and the worst was the class NLW.

Segmenting the patches increased 4.82%, 5.93%, and 4.45% the IoU metric for classes Crop, NLW, and BLW, respectively, concerning the IoU obtained where images were resized. Segmenting the patches obtained from the input images, without modifying the original size, may help to preserve significant features of the classes, then, the performance of the UNet-like model, under this scenario, is superior.

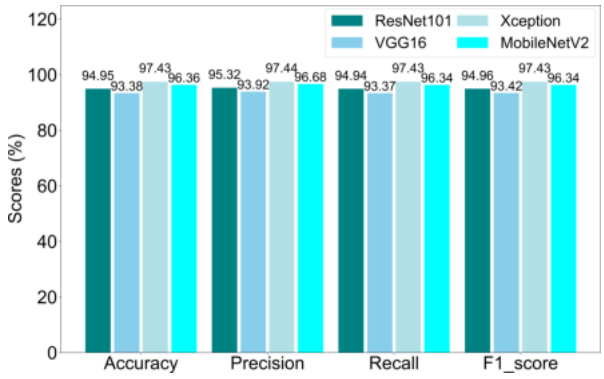

Fig. 8. Performance classification networks

In summary, when the images were resized during segmentation, the UNet-like model achieved a mean DSC of 84.27% and mIoU of 74.21%. In the other condition, when the images have been divided into patches, the UNet-like model achieved a mean DSC of 87.48% and a mIoU of 78.17%.

It is important to note that the magnitude values of the metrics used in our study are deemed acceptable as they surpass the performance of similar works reported in the literature. These works encompassed the segmentation of corn and weed plants in natural environments, as exemplified by the works of Quan et al. [16] and Picon et al. [15].

Additionally, our trained model can potentially segment other monocotyledon and dicotyledon plant species, given that the classes NLW and BLW, for which the architecture was trained, contain four species of each group with distinct growth stages. Moreover, the field variability was varied enough, making our trained model useful for segmenting a range of plant species.

In Figure 6 and Figure 7, we present two confusion matrices in which the performance of the UNet-like model can be appreciated. These matrices showcase the percentage of correctly and incorrectly classified pixels. In particular, Figure 6 shows the confusion matrix for the scenario where the input images were resized. In contrast, Figure 7 shows the confusion matrix for the scenario where the input images were divided into patches.

Under the two segmentation approaches, the classes Crop and BLW were better segmented

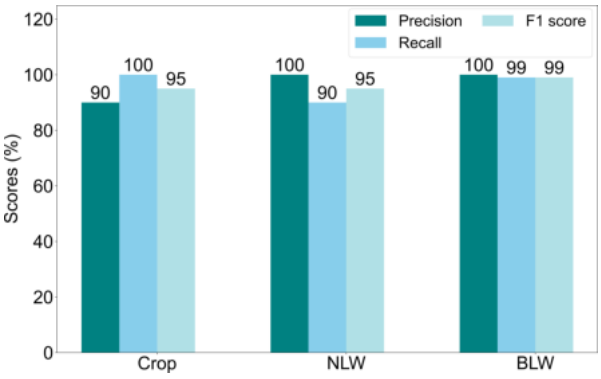

Fig. 9. MobileNetV2 classification performance

than the class NLW. In the first approach, the model was able to classify the pixels belonging to the Crop and BLW classes with a high degree of similarity, achieving an accuracy of 85%. However, the model's ability to classify the pixels of the NLW class was relatively lower, with an accuracy of only 61%.

In contrast, when the images were divided into patches, the pixels of the class BLW were better classified as such (91.0%), next the pixels of the class Crop (89.0%) and the worse classified were the pixels belonging to the class NLW.

In all the cases, the UNet-like model classified better the pixels belonging to the classes of plants into their corresponding class when the images were divided into patches, compared to that when they were solely resized.

It is also observed that the UNet-like model confused in more magnitude the pixels belonging to the classes of plants as if they were soil, under the two scenarios.

3.2 Performance of the Classification Networks

The implementation of transfer learning resulted in a notable improvement in classification performance. Specifically, the fully connected (FC) layers were tuned to our dataset to achieve this. The FC block comprised three layers, and it was observed that the classification accuracy was enhanced when the first two layers had 2,048 neurons.

Furthermore, the Adam optimizer with a learning rate of 0.0001 was utilized, and the categorical cross-entropy loss function was employed to minimize the error. The model was trained for 50 epochs on the complete dataset, with input images sized at 224 $\times$ 224 pixels for networks.

The macro performance of the networks ResNet101, VGG16, Xception, and MobileNetV2 on classifying the ROIs extracted from the segmented images are shown in Figure 8. It is worth mentioning that these metrics have been estimated under the segmentation scenario when the images were divided into patches.

As it is appreciated, Xception performed better, then MobiNetV2, and subsequently ResNet101, and the worse performance was depicted by VGG16, as the metrics Accuracy, Precision, Recall, and F_1-score indicate. In real-field applications, the inference time is crucial.

In this way, from the studied classification networks, the computation cost of MobileNetV2 network could be 8 to 9 times smaller than the rest of the architectures since it implements depthwise separable convolution (depthwise convolutions and pointwise convolutions), instead of conventional convolutions. Depthwise separable convolutions reduce trainable parameters [18].

For this reason, it was decided to present in Figure 9 the fine performance of MobileNetV2 model on classifying plants that belong to the classes Crop, NLW, and BLW. Analyzing first the metric Recall, it indicates that 100% of the images belonging to the class Crop were classified as such, 90% of the images of the class NLW were classified as such and 99% of the images from the class BLW were correctly classified by the MobileNetV2 model. Since the precision of the class Crop is 90%, it indicated that the model is misclassifying 10% of the plants of the class NLW as if they were corn, because the precision of this class, NLW, is 100%. Therefore, the metrics Precision, Recall, and F1-score make us realize that the better classified class was BLW. Finally, the mean classification performance among classes was 95%, 95% and 99%, for Crop, NLW, and BLW, accordingly, which is indicated by the F_1-score.

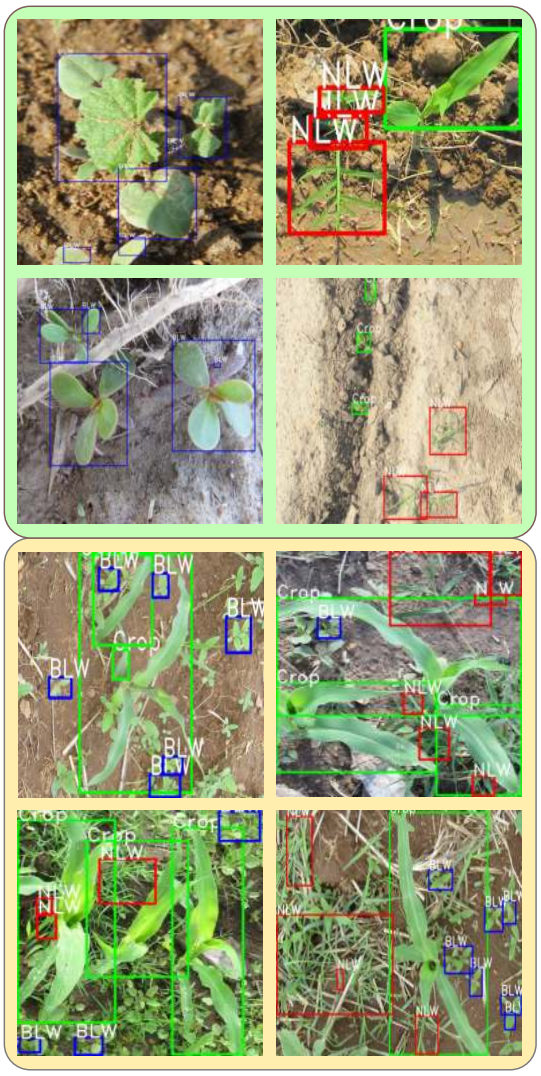

Fig. 10. Examples of the output images generated through the implemented detection method, utilizing both the segmentation and classification networks. The initial two rows exhibit images with low plant density, while the subsequent two rows showcase images featuring high plant density. Across all samples, the visual annotations include a green box denoting the crop, a red box indicating non-leaf weeds (NLW), and a blue box highlighting broadleaf weeds (BLW)

3.3 Detection Approach Visualization

Detecting objects in an image involves identifying the location and class of every object within the image. Figure 10 shows a sample of images in which the plant classes have been detected by applying our proposal. The first two rows of Figure 10 contain images that have a low density of plants, and occlusion of the foliage does not exist.

On the contrary, the images in the third and fourth rows show a high density of plants, and the foliage is partially covered in both cases. It's worth noting that the green boxes in all samples represent the Crop class, the red boxes represent the NLW class, and the blue boxes represent the BLW class. A visual inspection of the images with a low plant density indicates that almost all the green regions have been detected.

Nonetheless, since the localization of the plants is slightly related to the region provided by the segmentation model, more than one bounding box often appears in a simple image.

When high-density plant images are analyzed, it has been observed that most plant classes are accurately detected.

Nevertheless, due to the segmentation model's region extraction, it is common for multiple plants of the same classification to share a bounding box due to the density of foliage.

It is also appreciated that certain high-density plant images were not detected by the segmentation model due to the confusion of the pixels that belong to the plant classes with those of the soil. Although, in some cases, the detection covers part of the foliage of the plants, the implementation of this vision system for spraying herbicides under real corn fields is still adequate.

It is because the systemic herbicides are absorbed by the plants and gradually propagated throughout their vascular system, killing all their organs. Therefore, it has been observed that applying herbicides on a targeted section of plant foliage is adequate to eliminate them.

When the trained segmentation model considers multiple plants in a region, it could be tackled by subdividing the bounding box for spraying less area of the foliage.

4 Conclusion and Future Work

In this work, we present a method for detecting corn plants, as well as four narrowleaf (NLW) and four broadleaf (BLW) weed species in authentic corn fields. The proposed methodology comprises two distinct stages, namely segmentation and classification. A UNet-like architecture is employed from two different perspectives during the segmentation stage. The first consider segmenting the images entirely by resizing them, and the second approach consists of dividing the images into patches and then segmenting them.

In the classification stage, the four architectures ResNet101, VGG16, Xception, and MobileNetV2 have been evaluated on classifying the ROIs from the segmented images. Upon resizing the input images, the UNet-like model was able to attain a DSC of 84.27% and a mIoU of 74.21%. In the other scenario, when the images were divided into patches, the UNet-like model achieved a mean DSC of 87.48% and a mIoU of 78.17%. Regarding the classification networks, Xception performed better than MobiNetV2 and ResNet101. VGG16 showed the worst performance.

The segmentation model exhibited some limitations in accurately identifying the three classes of plants and the soil class. A significant proportion of pixels was frequently misclassified between these categories. Moreover, the models performed better in classifying the BLW class, but struggled with the NLW class, both in segmentation and classification. Notably, the models frequently mislabeled NLW as Crop.

In general, the models perform well despite the complexity of the dataset. In future work, we aim to enhance the segmentation performance of networks operating under high-density plants and develop a robust model capable of adapting to various field variabilities. Then, the dataset will be enlarged with more plant species and blur images captured with cameras mounted over moving agricultural tractors.

Acknowledgments

Francisco Garibaldi Márquez thanks to CONAHCYT for the scholarship granted.

References

1. **Bertels, J., Eelbode, T., Berman, M., Vandermeulen, D., Maes, F., Bisschops, R., Blaschko, M. B. (2019).** Optimizing the dice score and jaccard index for medical image segmentation: Theory and practice. Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 92–100. DOI: 10.1007/978-3-030-32245-8_11.
2. **Das, M., Bais, A. (2021).** DeepVeg: Deep learning model for segmentation of weed, canola, and canola flea beetle damage. IEEE Access, Vol. 9, pp. 119367–119380. DOI: 10.1109/access.2021.3108003.
3. **de Informacion Agroalimentaria y Pesquera, S. (2023).** Anuario estadístico de la producción agrícola. nube.siap.gob.mx/cierreagricola/.
4. **Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., Fei-Fei, L. (2009).** ImageNet: A large-scale hierarchical image database. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. DOI: 10.1109/cvpr.2009.5206848.
5. **Fawakherji, M. (2020).** Crop and weed classication using pixel-wise segmentation on ground and aerial images. International Journal of Robotic Computing, Vol. 2, No. 1, pp. 39–57. DOI: 10.35708/rc1869-126258.
6. **Gao, J., French, A. P., Pound, M. P., He, Y., Pridmore, T. P., Pieters, J. G. (2020).** Deep convolutional neural networks for image-based convolvulus sepium detection in sugar beet fields. Plant Methods, Vol. 16, No. 1. DOI: 10.1186/s13007-020-00570-z.
7. **Garibaldi-Márquez, F., Flores, G., Valentín-Coronado, L. M. (2023).** Segmentation and classification networks for corn/weed detection under excessive field variabilities. Proceedings of the Mexican Conference on Pattern Recognition, Vol. 13902, pp. 125–138. DOI: 10.1007/978-3-031-33783-3_12.

8. **Haralick, R. M., Shapiro, L. G. (1992).** Computer and robot vision. Addison-Wesley Publishing Company, Inc.
9. **Hashemi-Beni, L., Gebrehiwot, A., Karimoddini, A., Shahbazi, A., Dorbu, F. (2022).** Deep convolutional neural networks for weeds and crops discrimination from UAS imagery. Frontiers in Remote Sensing, Vol. 3, pp. 755939. DOI: 10.3389/frsen.2022.755939.
10. **Hu, C., Sapkota, B. B., Thomasson, J. A., Bagavathiannan, M. V. (2021).** Influence of image quality and light consistency on the performance of convolutional neural networks for weed mapping. Remote Sensing, Vol. 13, No. 11, pp. 2140. DOI: 10.3390/rs13112140.
11. **Hussain, N., Farooque, A., Schumann, A., McKenzie-Gopsill, A., Esau, T., Abbas, F., Acharya, B., Zaman, Q. (2020).** Design and development of a smart variable rate sprayer using deep learning. Remote Sensing, Vol. 12, No. 24, pp. 4091. DOI: 10.3390/rs12244091.
12. **Milioto, A., Lottes, P., Stachniss, C. (2018).** Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in CNNs. Proceedings IEEE International Conference on Robotics and Automation, pp. 2229–2235. DOI: 10.1109/icra.2018.8460962.
13. **Nedeljković, D., Knežević, S., Božić, D., Vrbničanin, S. (2021).** Critical time for weed removal in corn as influenced by planting pattern and PRE herbicides. Agriculture, Vol. 11, No. 7, pp. 587. DOI: 10.3390/agriculture11070587.
14. **Partel, V., Charan-Kakarla, S., Ampatzidis, Y. (2019).** Development and evaluation of a low-cost and smart technology for precision weed management utilizing artificial intelligence. Computers and Electronics in Agriculture, Vol. 157, pp. 339–350. DOI: 10.1016/j.compag.2018.12.048.
15. **Picon, A., San-Emeterio, M. G., Bereciartua-Perez, A., Klukas, C., Eggers, T., Navarra-Mestre, R. (2022).** Deep learning-based segmentation of multiple species of weeds and corn crop using synthetic and real image datasets. Computers and Electronics in Agriculture, Vol. 194, pp. 106719. DOI: 10.1016/j.compag.2022.106719.
16. **Quan, L., Wu, B., Mao, S., Yang, C., Li, H. (2021).** An instance segmentation-based method to obtain the leaf age and plant centre of weeds in complex field environments. Sensors, Vol. 21, No. 10, pp. 3389. DOI: 10.3390/s21103389.
17. **Ronneberger, O., Fischer, P., Brox, T. (2015).** U-Net: Convolutional networks for biomedical image segmentation. Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28.
18. **Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L. C. (2018).** MobileNetV2: Inverted residuals and linear bottlenecks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4510–4520. DOI: 10.1109/cvpr.2018.00474.
19. **Wang, A., Zhang, W., Wei, X. (2019).** A review on weed detection using ground-based machine vision and image preprocessing techniques. Computers and Electronics in Agriculture, Vol. 158, pp. 226–240. DOI: 10.1016/j.compag.2019.02.005.

Article received on 04/07/2023; accepted on 13/10/2023.
** Corresponding author is Luis M. Valentín-Coronado.*