Classification of Weeds and Crops at the Pixel-Level Using Convolutional Neural Neural Networks and Data Augmentation

Anderson Brilhador[1] , Matheus Gutoski, Leandro Takeshi Hattori,
Andrei de Souza Inácio, André Eugênio Lazzaretti, Heitor Silvério Lopes

Abstract— **The Pixel-Level Classification of crops and weeds is an open problem in computer vision. The use of agrochemicals is necessary for effective weed control, but one of the great challenges of precision agriculture is to reduce their use while maintaining high crop yields. Recently, automated weed control techniques based on computer vision were developed despite experiencing difficulties in creating agricultural datasets. One possible solution to the small volume of data available is Data Augmentation. This paper investigates the impact of individual data augmentation transformations on the pixel-level classification of crops and weeds when using a Deep Learning model. It also investigates the influence of input image resolution on the classification performance and proposes a patch augmentation strategy. Results have shown that applying individual transformations can be valuable to the model, but gets outperformed by the combination of all transformations. This work also finds that higher resolution inputs can increase the classification performance when combined with augmentation techniques, and that patch augmentation can be a valuable asset when dealing with a small number of high-resolution images. The method reaches the mark of 83.44% in Average Dice Similarity Coefficient, an increase of 19.96% percentage points compared to the non-augmented model.**

I. Introduction

The use of agrochemicals is necessary for effective weed control. However, they can cause negative impacts on the environment and affect human health [1]. Therefore, one of the great challenges of precision agriculture is to reduce the use of agrochemicals such as pesticides, herbicides, and fertilizers, while maintaining high crop yields. However, manually monitoring weeds and crops is a time-consuming activity, making it unfeasible for large-scale agriculture.

The development of Automated Weed Control (AWC) techniques is critical to reduce the time and cost of this task. These techniques should be able to locate and recognize weed species in a crop, allowing the selective application of agrochemicals or the mechanical extraction of weeds. They can be incorporated into robots, unmanned aerial vehicles (UAVs), tractors, and other agricultural equipment.

Pixel-Level Classification (PLC) is a technique based on Computer Vision that allows to locate and segment certain parts of an image. In the context of AWC, it has been used to identify weeds and other plants [1], [2], [3].

In recent years, Deep Learning (DL) methods have been successfully used to solve a wide array of problems related to computer vision [4]. However, one of the main drawbacks of DL methods is the need of large datasets to achieve good performance and generalization [5], [6].

Notwithstanding, the process of acquiring images in the agricultural domain is not trivial. It requires access to cultivation sites, synchronization of the image acquisition with the growth of the plants, and expert assistance to annotate the data acquired [7]. Due to these difficulties, annotated datasets of this area usually have an insufficient amount of samples to be used with DL methods [7], [8], [9]. Under these circumstances, the use of Data Augmentation (DA) techniques is essential for the development of weed control methods based on DL. Recent works of pixel classification applied to weed and crop images have shown that DA can improve the performance of their models [3], [10]. However, these studies do not explore the impact of each transformation applied to images, using DA only as a replication technique to improve the results.

This work investigates the application of different spatial data augmentation transformations in a small set of images. The transformed data is fed to a Convolutional Neural Network (CNN) that performs pixel-level classification. The main contributions of our work are:

- Investigation of the impact on the performance of pixel-level classification of crops and weeds using a CNN with DA;
- Investigation of the influence of image resolution on PLC using different DA techniques;
- Proposition of the patch augmentation for the PLC problem.

II. Related Work

Several recent works have used CNNs in the context of crop and weed pixel classification, as follows.

In [11], authors use the SegNet architecture [12] for pixel-level classification of sugar beet plants and weed. The network was trained with 3-channel images that include the NIR channel, the Red channel of the RGB image, and the NDVI map. A similar approach was presented in [2], in which a modification of the SegNet architecture was proposed. The network takes as input 14-channel images that include data from various vegetation indexes.

On the other hand, a solution using a Fully Convolutional Network (FCN) combined with an encoder-decoder structure is proposed in [13], where a sequence of images fed the model with spatial information. The use of FCNs for crop and weed segmentation was also explored in [1].

[1]Brilhador, Anderson. brilhador@utfpr.edu.br
All authors are with Federal University of Technology - Paraná, Av. Sete de Setembro, 3165, Curitiba, Paraná, Brazil

In [3], authors accurately classified crops and weeds using a sequence of two convolutional neural networks (CNNs) applied to RGB images. The first step uses a UNet-based network [14] and aims at separating pixels between two classes: soil and plant. After that, each plant is classified as crop or weed using a fine-tuned VGG-16 model [15].

Due to the lack of sufficiently large datasets, the authors in [16] generated synthetic data to train the SegNet model. The augmentation algorithm randomly creates the main features of the target environment, such as crop and weed species, soil type and light conditions.

The use of data augmentation has been studied in different contexts to explore its ability to increase the performance of DL models. For instance, a new method of data augmentation to prevent overfitting in an Acoustic Event Recognition (AER) task is presented in [17]. Experimental results have shown improvements over the state-of-the-art.

Data augmentation in different CNN architectures was studied in [18] for classifying soft biometric attributes. New samples were generated using small random transformations, such as rotation, cut, swirl, vertical rotation, horizontal rotation, salt and pepper, and Gaussian noise.

Similar approaches to classify images were used by [5], [19]. The authors have also shown the use of Generative Adversarial Networks (GAN) [20] for the data augmentation task.

Several works explored the use of data augmentation to improve the generalization ability of classifiers, although they seldom evaluate the individual effect of each transformation.

This work differs from previous works because it investigates the effect of data augmentation for pixel-level classification of crops and weeds in a more detailed way. Our experiments are designed to evaluate each transformation individually, as well as combined. Moreover, we propose a patch augmentation strategy that consists of breaking high-resolution images into smaller pieces, which is unprecedented in crop and weed pixel-level classification research.

III. Methods

A. Overview

Figure 1 presents a block diagram of the proposed method, and the following Sections provide detailed explanations. The process goes as follows: first, the raw dataset goes through a preprocessing step, as described in Section III-B. Then, the data is split in a 5-fold cross-validation procedure, as described in Section III-C. The next step is to augment the folds reserved for training using the augmentation strategies presented in Section III-D. The modified U-Net architecture, presented in Section III-E, receives the augmented data. We implement the data augmentation in an on-line manner, except for the patch augmentation strategy, implying that each epoch of training receives previously unseen samples. Once the model is trained, test samples are classified, and evaluation metrics are computed, as shown in Section III-F. The patch augmentation strategy requires one extra step, which is generating patches of the test images. All patches are treated as individual images.

B. Preprocessing

The first pre-processing step applied to the data is a [-1,1] normalization in the pixel values. The second step is to reduce the image resolutions since the original dataset contains high-resolution images which are not practical to work with due to hardware constraints. We consider the input sizes of 96×96, 224×224 and 448×448 pixels.

C. Training and Data Splits

Training and test splits were generated using a 5-fold cross-validation procedure. Data augmentation was applied only to the training folds, while the test folds remained unchanged. This strategy allows the analyzes of the generalization capability of the model over a better sampling policy than a simple hold-out procedure.

The patch augmentation strategy crops both training and test folds. In this case, the evaluation considers all patches generated by the algorithm. The number of images generated for each fold after using data augmentation is shown in Table I.

The training process proceeds until the accuracy stagnates for ten consecutive epochs. The Adadelta optimizer was used with an initial learning rate of 1 with a decay factor of 0.95. Dropout layers have a drop rate of 10%.

TABLE I: Number of images on each subset of data after the application of data augmentation techniques.

Subset	**CFWID Raw**	**Augmented All**	**Patch Augmented** 96×96	224×224	448×448
Training	48	1152	7392	1440	432
Test	12	12	1848	360	108
Total	60	1164	9420	1800	540

D. Dataset and Data Augmentation

This work uses the Crop/Weed Field Image dataset (CWFID) [7] that contains pixel-level annotations of carrot and weed images. A total of 60 images with resolution of 1296×966 pixels were captured under controlled lighting conditions by an autonomous agricultural robot. Table II presents examples of images with the spatial transformations proposed in this work and their respective ground truth.

There are different DA strategies to improve the performance of classifiers. This work was limited to spatial transformations (horizontal flip, vertical flip, rotation, width shift, height shift, shear, and zoom). In addition to these transformations, we apply a patch augmentation strategy, which leaded to interesting results in recent works in other contexts [21], [22].

The patch augmentation strategy consists in applying a $n \times n$ grid over the original image and its corresponding ground-truth. The image is then divided into small pieces, which are treated individually. A padding procedure was used to fill in the borders of the original image with zeros to ensure equally sized patches. Figure 2 illustrates the patch augmentation strategy.

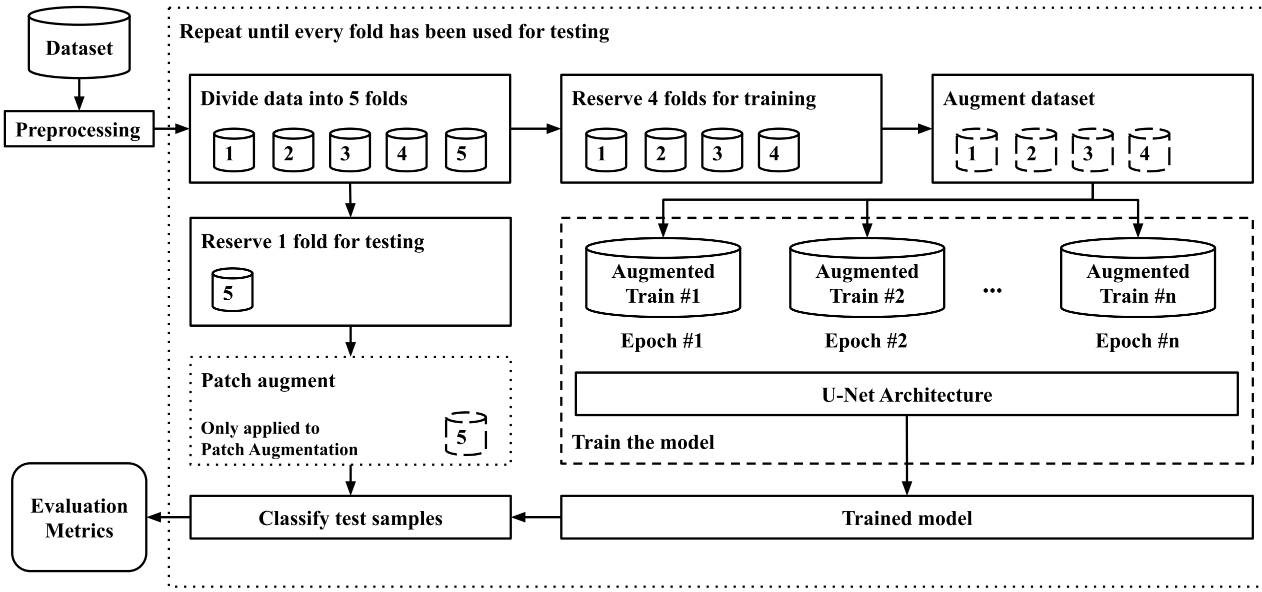


Fig. 1: Overview of the proposed method.

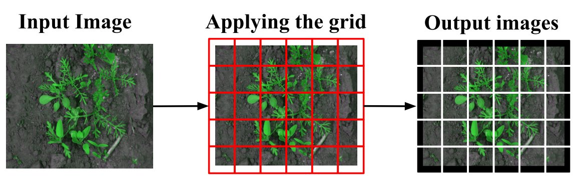


Fig. 2: Illustration of the patch augmentation strategy.

E. *Modified U-Net Architecture*

Pixel-Level Classification (PLC) is the problem of assigning a label (class) to each pixel of a given image. PLC requires a supervised learning approach, and, therefore, it is needed a set of previously labeled data. PLC is also known as semantic segmentation or pixel-wise segmentation. Amongst the many methods for PLC, this work uses a CNN architecture.

U-Net is a DL model proposed by [14] for an image segmentation problem. Its architecture consists of an encoding network and a decoding network. The encoding network follows the typical architecture of a CNN, composed of a series of convolutions, Rectified Linear Units (ReLU) and max-pooling operations. Layers are stacked so that each layer learns features of increasing levels of complexity while simultaneously performing downsampling. On the other hand, the decoding network progressively increases the resolution of the learned features. The model concatenates the high-level representations of the encoding network and the upsampled outputs of the decoding network to combine global information with localization accuracy.

In our implementation, the model was incremented with a convolutional layer with a kernel size of 1×1 to obtain the desired number of outputs. The remaining convolutional layers have a kernel size of 3×3. Each convolution is followed by a ReLU activation function, except for the last layer, which performs classification using a softmax function. The rest of the model employs max-pooling and transposed convolution operations with 2×2 kernels. The final network output is a pixel-level mask that shows the class of each pixel. For our PLC problem, there are three possible classes for each pixel: soil, weed, or crop.

Besides the above-mentioned changes, other adjustments to the architecture were done (see Figure 3). The first adjustment was the input size of the network, since it needs to match the input sizes used in this work. The second adjustment was replacing all ReLU activation functions with the Exponential Linear Unit (ELU). The ELU activation functions tends to converge faster and provides more accurate results [23]. Also, we optimize the model using the Adadelta algorithm instead of the traditional Stochastic Gradient Descent (SGD). Finally, as proposed by [24], dropout layers were placed in between convolutional layers to reduce overfitting.

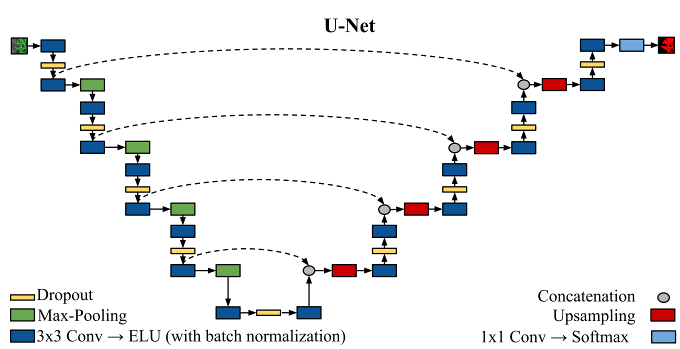
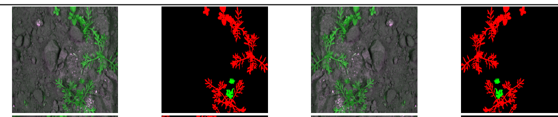
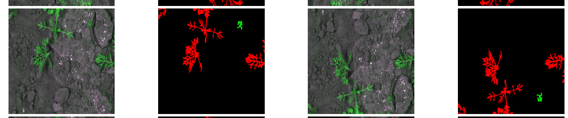
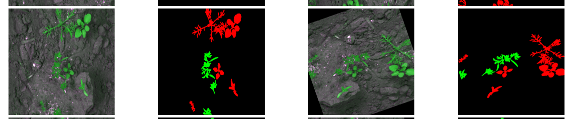
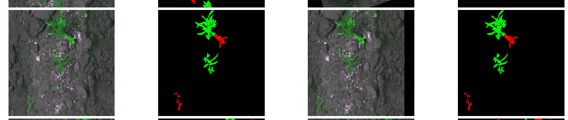
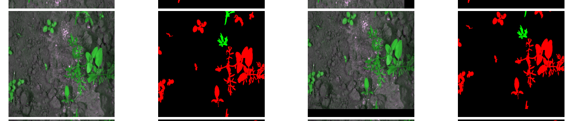
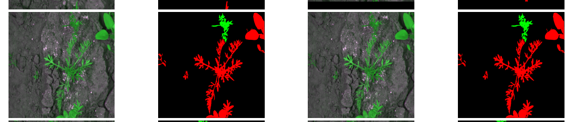
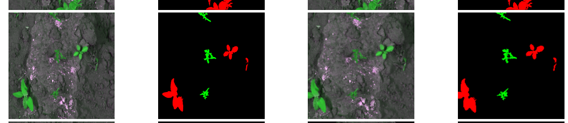
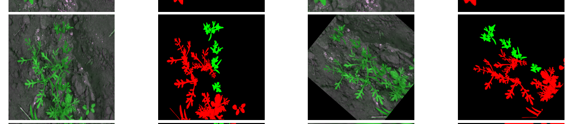
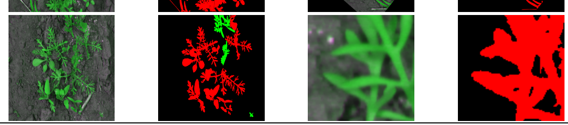


Fig. 3: Our version of the U-Net architecture. Figure adapted from [14].

TABLE II: Examples of dataset images and transformations applied in data augmentation. The pixels in black, red and green represent the soil, weeds and crops, respectively.

Transformation	**Description**	**Policy**	**Examples**			
			Raw		**Transformed image**	
			Image	**Ground truth**	**Image**	**Ground truth**
Horizontal flip	Randomly invert images horizontally	-				
Vertical flip	Randomly invert images vertically	-				
Rotation	Random rotations in a given range of degrees	0 to 90 degress				
Width shift	Shifts the image horizontally randomly within a fraction boundary	1/10 maximum fraction				
Height shift	Shifts the image vertically randomly within a fraction boundary	1/10 maximum fraction				
Shear	Shear in the image counter-clockwise within a maximum degree limit	0 to 2 degrees				
Zoom	Randomly generated zooms within a scale	-0.2 to 0.2 scale				
All	All randomly applied transformations	Policies are preserved				
Patch augmentation	Divide the image into smaller parts	Determined resolution				

F. Evaluation Metrics

The Dice Similarity Coefficient (DSC) [25] was used to analyze the outcome of the experiments. It is a measure of spatial overlap that ranges from 0 (no spatial matching between two sets of pixels) to 1 (complete overlap). In the literature, DSC has been used for several purposes, including the comparison between segmentation methods. The average DSC (Avg. DSC) is also computed between all classes to provide a general analysis, ensuring that all classes have the same weight. This is necessary because the dataset is heavily unbalanced. Equation 1 presents the DSC for binary labels:

$$\mathrm{DSC}_i = \frac{2TP_i}{2TP_i + FP_i + FN_i}, \tag{1}$$

where TP_i, FP_i and FN_i are number of true positives, false positives, and false negatives for class i. Finally, we multiplied all the metrics by 100 to convert them into percentages.

The average DSC is the arithmetic mean of the DSC of each class, as defined by:

$$\text{Avg. DSC} = \frac{1}{C}\sum_{i=1}^{C} DSC_i, \tag{2}$$

where C is the number of classes.

IV. Experimental Results and Discussion

The experiments were run on a workstation with Intel Core-i7 8700 processor, 32GBytes RAM, Nvidia Titan Xp GPU, and the Ubuntu 18.04 LTS operating system. The Keras[1] framework with the TensorFlow backend were used for the development of the DA transformations and the U-net.

A. Experiments

The first set of experiments were designed to verify what kind of DA could provide the best classification performance

[1] `https://keras.io/`

in the task of crop and weed pixel classification. Hence, the complete workflow (shown in Section III) was conducted using one kind of DA strategy at a time, including patch augmentation.

Besides the evaluation of individual DA transformations in the images, we also performed experiments combining all types of DA (excluding patch augmentation). A baseline result was generated using the raw dataset (without DA) for comparison. We conduct this experiment with 96×96 images.

The second set of experiments aim at analyzing how the input image size affects the final classification performance. We conduct the same experiments described above with two additional input sizes: 224×224, and 448×448 pixels.

B. Results and Discussion

The experimental results of this work are summarized in Table III. In this Table, rows compare the several DA transformations against the baseline (no DA), whilst main columns compare the image resolutions for all classes.

Regarding the individual impact of each type of DA, we observed that the patch augmentation strategy provided the best results in terms of Avg. DSC for images with 96×96 pixels, with an increase of 11.3% percentage points when compared to the baseline (raw dataset). However, the performance of patch augmentation falls off as the input resolution increases, as seen for the 224×224 and 448×448 sets. Possibly the performance loss is due to decrease of the number of patches, as the resolution increases (shown in Table I). This suggests that patch augmentation works better with small patch sizes. Another important point about patch augmentation is that it preserves the original information, since it is not affected by resizing algorithms.

Table III also shows that all DA transformations led to increased performance over the baseline. Particularly, the shear transformation was overall slightly worse than other augmentation strategies, which presented similar results. The combination of all spatial transformations achieved the best result for 448×448 pixels images, and a close second for 224×224 images. These results suggest that a combination of many types of transformations may be valuable for image segmentation tasks. This strategy reached 83.44% in the Avg. DSC, the overall best result of all experiments. This corresponds to an increase of 19.96% percentage points relative to the model trained using only raw images.

Regarding the second experiment, Table III shows that as the resolution increases, DA strategies become more effective, except for patch augmentation for the reason discussed above. However, using raw dataset (no DA transformation), in its turn, led to inverse results, i.e., performance decreases as resolution increases. Generally, high dimensional data require larger datasets to achieve good performance, which may explain the reason for the performance loss.

Overall, it can be observed that the model easily discriminates between soil and other classes. In one hand, this is good considering the semantic level of the image segmentation problem in hand. On the other hand, this is expected, since the soil color and texture are quite different than weeds and crops. The real challenge, as seen in the Table, is to discriminate between weeds and crops, which have very similar visual attributes. Despite the difficulties, in the best cases our model achieved 87.02% DSC in the weed class and 66.31% in the crop class.

V. Conclusions

This paper investigated the impact of different data augmentation techniques on the performance of pixel-level classification of crops and weeds. Three main points were investigated: (1) how individual data augmentation techniques and their combination affect the classification performance; (2) what is the impact of increasing the input image resolution; and (3) what is the performance of the proposed patch augmentation technique concerning other spatial transformations.

Regarding the first point, our experiments have shown that all types of spatial transformations used in this work boost the performance of the classifier, showing significant gains over the baseline. Actually, the combination of all transformations ultimately achieved the highest performance in terms of Avg. DSC. This result suggests that introducing variability, as opposed to solely increasing the volume of data using single transformations, is an important factor for the difficult task of crop and weed classification.

The second point was to investigate the impact of the image resolution. Results suggest that, on one hand, a higher resolution input is beneficial in cases where the number of images increases. For instance, experiments that considerably increased the number of images on the dataset using spatial transformations have shown a performance gain in higher resolution settings. On the other hand, increasing the input resolution was not beneficial in cases where the number of images decreased or remained constant. The patch augmentation and the raw dataset baseline suffered performance losses as the input resolution increased. The curse of dimensionality may be the explanation for this result since experience indicates that datasets of high-resolution images generally require a large volume of training data.

The third point concerns the patch augmentation technique proposed in this work. The results showed that this technique works particularly well with small patch sizes since it considerably increases the volume of data. This increase is not only due to the total number of images generated but, also, because the number of pixels in the dataset remains the same, i.e. the information is preserved. Other data augmentation transformations lose information since they suffer resizing operations. Finally, we conclude that the patch augmentation strategy can be a valuable approach when images are scarce and have high resolution.

Future works will investigate new data augmentation techniques and ways to combine patch augmentation with other types of transformations, as well as the use of smaller patch sizes and strides. It is ultimately expected to create a large volume of high diversity images from small-sized samples,

TABLE III: PLC performance using the U-NET network with different DA transformations and images with different resolutions: 96×96, 224×224 and 448×448.

Data Augmentation	Image Resolution 96×96				Image Resolution 224×224				Image Resolution 448×448			
	Avg.[%]	DSC[%]			Avg.[%]	DSC[%]			Avg.[%]	DSC[%]		
		Soil	Weeds	Crops		Soil	Weeds	Crops		Soil	Weeds	Crops
Raw dataset (no DA)	71.08	99.14	76.56	37.54	65.89	99.04	76.04	22.59	63.48	98.84	74.95	16.65
Flip vertical and horizontal	80.09	99.30	83.21	57.77	**83.23**	99.25	84.58	**65.86**	81.33	99.29	83.23	61.47
Rotation max 90 degrees	81.40	99.36	83.92	**60.91**	82.70	99.31	84.75	64.03	83.40	99.32	**85.15**	65.74
Shift height and width	78.17	99.28	81.73	53.51	79.37	99.24	82.57	56.30	81.57	99.27	83.06	62.38
Shear	76.29	99.23	79.56	50.09	80.53	99.23	82.75	59.60	80.70	99.26	82.88	59.94
Zoom	78.01	99.27	81.85	52.90	80.17	99.27	82.34	58.90	81.15	99.32	83.45	60.70
All spatial transformations	79.82	99.33	83.89	56.23	83.17	99.37	85.36	64.80	**83.44**	99.33	84.69	**66.31**
Patch augmentation	**82.38**	**99.60**	**87.02**	60.53	82.25	**99.59**	**86.42**	60.75	74.39	**99.61**	81.22	42.34

and so improve the quality of results and generality of the method for difficult segmentation tasks.

Acknowledgements

A. Brilhador thanks for support from the Federal Technological University of Paraná; M. Gutoski thanks CNPq for the scholarship number 141983/2018-3; L. T. Hattori thanks CAPES for the scholarship; A. S. Inácio thanks for support from the Federal Institute of Santa Catarina; H.S.Lopes thanks CNPq for the research grants no. 311778/2016-0 and 423872/2016-8; Special thanks to NVIDIA Corporation for the donation of the Titan-Xp GPUs used in this work.

References

[1] P. Lottes, J. Behley, N. Chebrolu, A. Milioto, and C. Stachniss, "Joint stem detection and crop-weed classification for plant-specific treatment in precision farming," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 8233–8238.

[2] A. Milioto, P. Lottes, and C. Stachniss, "Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in cnns," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 2229–2235.

[3] M. Fawakherji, A. Youssef, D. Bloisi, A. Pretto, and D. Nardi, "Crop and weeds classification for precision agriculture using context-independent pixel-wise segmentation," in *Third IEEE International Conference on Robotic Computing (IRC)*. IEEE, 2019, pp. 146–152.

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.

[5] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," *arXiv preprint arXiv:1712.04621*, 2017.

[6] J. Ferreira, M. Ferro, B. Fernandes, M. Valenca, C. Bastos-Filho, and P. Barros, "Extreme learning machine autoencoder for data augmentation," in *IEEE Latin American Conference on Computational Intelligence (LA-CCI)*. IEEE, 2017, pp. 1–6.

[7] S. Haug and J. Ostermann, "A crop/weed field image dataset for the evaluation of computer vision based precision agriculture tasks," in *Computer Vision - ECCV 2014 Workshops*, 2015, pp. 105–116. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-16220-1_8

[8] A. Brilhador, D. A. Serrarens, and F. M. Lopes, "A computer vision approach for automatic measurement of the inter-plant spacing," in *Iberoamerican Congress on Pattern Recognition*. New York: Springer, 2015, pp. 219–227.

[9] N. Chebrolu, P. Lottes, A. Schaefer, W. Winterhalter, W. Burgard, and C. Stachniss, "Agricultural robot dataset for plant classification, localization and mapping on sugar beet fields," *The International Journal of Robotics Research*, 2017.

[10] I. Sa, M. Popović, R. Khanna, Z. Chen, P. Lottes, F. Liebisch, J. Nieto, C. Stachniss, A. Walter, and R. Siegwart, "Weedmap: a large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming," *Remote Sensing*, vol. 10, no. 9, p. 1423, 2018.

[11] I. Sa, Z. Chen, M. Popović, R. Khanna, F. Liebisch, J. Nieto, and R. Siegwart, "Weednet: Dense semantic weed classification using multispectral images and mav for smart farming," *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 588–595, 2017.

[12] V. Badrinarayanan, A. Handa, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling," *arXiv preprint arXiv:1505.07293*, 2015.

[13] P. Lottes, J. Behley, A. Milioto, and C. Stachniss, "Fully convolutional networks with sequential information for robust crop and weed detection in precision farming," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 2870–2877, 2018.

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. New York: Springer, 2015, pp. 234–241.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[16] M. Di Cicco, C. Potena, G. Grisetti, and A. Pretto, "Automatic model based dataset generation for fast and accurate crop and weeds detection," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 5188–5195.

[17] N. Takahashi, M. Gygli, B. Pfister, and L. Van Gool, "Deep convolutional neural networks and data augmentation for acoustic event detection," *arXiv preprint arXiv:1604.07160*, 2016.

[18] M. R. Aquino, M. Gutoski, L. T. Hattori, and H. S. Lopes, "The Effect of Data Augmentation on the Performance of Convolutional Neural Networks," in *Proceedings of 4th IEEE Latin American Conference on Computational Intelligence LA-CCI*. Piscataway, NJ: IEEE Press, 2017, pp. 1–6.

[19] J. Shijie, W. Ping, J. Peiyi, and H. Siping, "Research on data augmentation for image classification based on convolution neural networks," in *Chinese Automation Congress (CAC)*. IEEE, 2017, pp. 4165–4170.

[20] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[21] L. T. Hattori, M. Gutoski, M. R. Aquino, and H. S. Lopes, "Patch-Based Convolutional Neural Network for the Writer Classification Problem in Music Score Images," in *Proceedings of XIII Brazilian Congress on Computational Intelligence*. Piscataway, NJ: IEEE Press, 2017, pp. 1–6.

[22] R. Takahashi, T. Matsubara, and K. Uehara, "Data augmentation using random image cropping and patching for deep cnns," *arXiv preprint arXiv:1811.09030*, 2018.

[23] D. A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," *arXiv preprint arXiv:1511.07289*, 2015.

[24] J. Zhuang, "Laddernet: Multi-path networks based on u-net for medical image segmentation," *arXiv preprint arXiv:1810.07810*, 2018.

[25] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, pp. 297–302, 1945.