

# Recognition of Weeds in Cornfields

Gábor Hartyányi

ORCID ID: 0009-0002-0524-0495

Image Processing Research Laboratory  
 University of Pannonia, 8200 Veszprém  
 Egyetem u. 10, Hungary  
 also with Axial Ltd, 6500 Baja  
 Email: hartyanyi.gabor@gmail.com

László Czúni

ORCID ID: 0000-0001-7667-9513

Image Processing Research Laboratory  
 University of Pannonia, 8200 Veszprém  
 Egyetem u. 10, Hungary  
 Email: czuni.laszlo@mik.uni-pannon.hu

**Abstract**—In terms of weed control, existing precision spraying solutions seek to reduce the unwanted impact of spraying by separate field scanning from mostly birds’ eye view. In our study, we propose a hybrid approach in which the mechanical hoeing and the spraying is done simultaneously accomplished by weed recognition from a lower position where the plants’ leaves do not cover weeds. We demonstrate the line and the weed recognition methods on a dataset collected from corn fields and compare different convolutional neural networks. We also investigate the feasibility on two widely known embedded platforms.

## I. INTRODUCTION

WHILE the population of Europe and world is still increasing the possible incorporation of new lands to agriculture is very limited. Weed is a major factor to limit crop yields, beside their physical clearing, spraying is a main general solution to this problem. However, the EU Green Deal agreement [1] aims 50% reduction of applied chemicals, thus conventional spraying techniques should be revised. In traditional, large field applications whole areas are being sprayed resulting in wasted chemicals on healthy and intact plants and on bare soil. The utilization of machine vision techniques for the detection of weed has been a target for decades, a good overview of different approaches can be found in [2]. Computer vision techniques have to face lots of problems if applied on the field. The leaves of weeds and crops often overlap each other at late growth stages, especially if images are taken from above, making them indistinguishable. Additionally, the plant’s leaves may be obscured or damaged by unwanted material including dead leaves or clay, making identification difficult. Maize is the most produced grain in the world, with more than 1.2 billion tons produced in 2021. In Europe, the area affected by corn cultivation was approximately 20 million hectares [3]. Thus, the development of solutions for corn alone can have a significant impact on environmental protection, and the various techniques can be adapted to other crops as well. The main contributions of our article are: We are proposing a hybrid approach combining hoeing and spraying. Cameras, fixed to the cultivator, are to capture the areas between the stems of corn (or between

We acknowledge the financial support of the Hungarian Scientific Research Fund grant OTKA K-135729. We are grateful to the NVIDIA corporation for supporting our research with GPUs obtained by the NVIDIA Hardware Grant Program.

lines); convolutional neural networks (CNNs) recognize weeds and spraying is concentrated only on those areas near stems. Hoeing is made between the lines simultaneously, there is no need for multiple scanning of the fields. Fig. 1 illustrates the region of interest (ROI) areas for possible spraying. We introduce a new free annotated dataset of images of corn lines. Binary labels indicate the presence of weeds. We investigate the use of a popular scalable DNN (EfficientNet [18]) and less complex CNNs for weed recognition. The feasibility on micro-controllers is also part of our study. To narrow down the target area for spraying the physical setup is calibrated with image homography.



Fig. 1: Corn fields being hoed and only the ROI areas to be sprayed where weed is detected.

## II. LITERATURE OVERVIEW

Traditional approaches typically consist of four main steps: pre-processing, segmentation, feature extraction, and classification. Pre-processing tries to "standardize" global image properties, while segmentation is to separate vegetation from background. Since weed and crops have similar properties it is difficult to find the proper features and their representation to achieve the best possible classification. Four feature categories can be identified in papers, such as spectral features (mean and standard deviation of RGB, HSV, and chlorophyll vegetation index values), textural features, morphological features, and spatial contexts.

For textural features common techniques can be utilized, for example in [4] single-level Haar discrete wavelet decomposition was used to obtain four sub-images (approximate image,

vertical details, horizontal details, and diagonal details), and then gray level co-occurrence matrices were extracted from these sub-images for weed detection. In [5] Gabor filters were applied but also the co-occurrence matrix features were finally calculated. The shape of leaves can be very characteristic for recognition and dozens of such traditional descriptors have already been utilized for weed recognition: eccentricity, circularity, convexity, elongatedness, invariant moments, just to mention a few. The above mentioned approaches did not consider the spatial context, while the sowing pattern of crops is typically very specific: they are sowed or planted in almost straight lines thus the spatial contexts or position information could help to improve the recognition process. It is natural to think of the variants of the Hough transformation, linear regression, the vanishing point, or the frequency analysis of the lines or repetitive patterns. More sophisticated approaches (such as [6] which uses dynamic programming and energy optimization) can also handle curved lines. However, applying strict assumptions about crop positions can result in false detections. In [7] the upper limit of detection accuracy was investigated when using information about sowing geometry and positions. The uncertainty in real crop positions and the disturbing effect of weeds can have a significant effect on detection accuracy, thus complex solutions are required.

Beside RGB cameras, special sensors such as depth cameras can also be used for weed recognition. For example in [8], beside color, position, and texture features also depth features, obtained by a special RGB-D camera, were also utilized to recognize weeds in wheat fields. The AdaBoost algorithm was employed for the integrated learning of multiple classifiers. Experimental results showed accuracy between 81 and 88%, depending on the growth phase of wheat (the different experiments used 50-600 images).

Considering the theoretical and practical problems of the above specified four main steps, there is no surprise for the breakthrough of deep learning methods. As an early attempt to overcome the weak generalization ability of manually designed features [9] used K-means clustering to construct a feature dictionary, fed to a single-layer network, to create an identification model. The approach in [10] can be considered as a hybrid solution where both hand crafted features (requiring segmentation) and DNN features, generated by a pre-trained GoogLeNet [11] network, were used with four kinds of clustering methods. Interestingly, the technique was used to cluster four kinds of weeds but the number of test images were much below a thousand. In [12] an embedded system on a UAV was introduced utilizing the YOLOV3-tiny network to detect the pixel coordinates of weeds in images. The mean Average Precision (mAP) was 72.5% at 2FPS on a mobile device. The average positioning error was 10.31 cm. Tests were carried out on a total of 2000 images, taken at 2 m high, of winter wheat with 5 types of weeds. The most relevant paper from our point of view is [13] where a classification approach of *Zea mays* L. (corn), narrow-leaf weeds, and broadleaf weeds from multi-plant images are presented. Compared to previously discussed articles, a large image dataset was generated: 13,000

recordings were made in natural field conditions, at different locations and at different stages of plant growth. The ROIs were detected using connected component analysis, whereas the classification was based on VGG [14] and Xception [15] CNNs (and alternatively by SVMs). The best method for weed classification, at early stages of growth and in natural corn field environments, was the CNN-based approach, as indicated by the 97% accuracy obtained.

For the reader interested in hand-designed feature methods we propose to read the review of Wang et al. in [2] while for more recent DNN approaches go for [16].

### III. A MAIZE IMAGE DATASET

Contrary to hand-designed approaches machine learning methods, especially DNNs, don't require much pre-processing but large datasets with enough generality are a must. While some articles were trying to recognize the different types of weeds (e.g. [10]) or tried to increase the variety of viewpoints (f.e. [13]) we have different purposes: Since between the lines of corn hoeing is made, weeds are to be detected (and sprayed if found) only between maize stems. Weed types are out of interest and three types of images are to be classified: only weed, only maize, and weed and maize. The cameras can be placed on the cultivators approximately 25 cm high and 25 cm laterally from the corn row, with the optical axis of 45 degrees to the ground plane.



Fig. 2: Example images of the "Corn and Weed" dataset. Top: clear maize. Middle: only weeds. Bottom: maize and weeds.

Images of our publicly available dataset (downloadable at <https://keplab.mik.uni-pannon.hu/images/caw/>) were made in a second-sown corn field. Sowing time was late May - early June 2022. The average row spacing was usually 70 cm, while the distance between the stems was on average 25 cm. The height of the plants varied between 20-50 cm depending on the nutrient and water supply of the area. The shots were made with a GoPro 7 camera at  $2704 \times 1520$  resolution and at an average speed of 4 km/h, with different corn line orientations. The dataset contains 816 images with only weed, 1231 images with only corn, and 1796 images with corn and weed. Original images are downscaled to  $640 \times 480$ , example photos are in Fig. 2.

#### IV. DETECTION OF CORN LINES AND SAFETY MARGIN AREA

For the most accurate localization of the ROI we make the following steps: First we segmented corn stems. For the instance segmentation of corn stems we used Mask-RCNN [17] pre-trained on the COCO dataset. For transfer learning with two classes (background and corn stem) 50 images (with circa 200 corn plants) were manually annotated. We used stem bottom endings as the lowest points of Mask-RCNN masks to fit lines with linear regression. ROI was set with planar homography (see Subsection IV-A).



Fig. 3: Segmented maize stems with Mask R-CNN.

##### A. Planar Homography for ROI Designation

To find the border lines of the ROI the size of the safety margin should be considered. In our layout 10 cms were given on both sides of corn lines. To determine the border lines in the image space we computed the homography matrix with the help of ArUco markers. By applying plain homography we assumed the smoothness of the ground (the relative pose of the camera plane and soil at the stem endings is constant). Naturally, this is not always true but considering the spread of the spray we accepted the resulting inaccuracy. The result is illustrated in Fig. 4.

Starting from our initial dataset now we arrived to a smaller set: there are only two labels (weed free and with weed) and to avoid a very unbalanced configuration the number of weed free images were limited. Tab. I gives the number of images per category in our experiments.



Fig. 4: ROI defined by stem endings and homography of safety margins.

TABLE I: The ROI based dataset used in experiments.

	Weed free	With weed	Total	Percentage
<b>Training images</b>	851	808	1659	72%
<b>Validation images</b>	230	209	439	19%
<b>Test images</b>	59	160	219	9%
<b>Total</b>	1140	1177	2317	100%

#### V. COMPARISON OF DIFFERENT WEED RECOGNITION MODELS

Assuming approximately 15 km/h average speed of the cultivator, circa  $70^\circ$  viewing angle, and 10% overlapping of images at least 4 FPS processing speed should be reached. There are two main purposes of the following experiments: First, to investigate the effect of masking: what happens if the whole area (i.e. the context) is considered during the classification at the ROI. Second, to find the limit to minimize the complexity of the applied CNNs so to increase the processing speed without a painful degradation of accuracy. In 2019, Google Brain published the open source EfficientNet [18] network family for image classification. The members of the family are the differently scaled versions (from B0 to B7) of the base model, B7 being the largest variant achieving state-of-the-art Top-1 accuracy on ImageNet in 2019. It was created with a compound scaling method to scale the depth (number of layers), the width (number of kernels in a layer), and resolution (size of input image) of an existing model and a baseline network with fine-tuned layers, in a balanced manner, to consider the computation limits. We used the ImageNet pre-trained B0 version without the top classification parts after adding two dense hidden layers with 512 and 128 neurons and two output neurons.

Tab. III compares results showing almost perfect classification accuracy on both masked and whole area images. Thus our next step was to create CNNs with decreasing number of parameters to reach the smallest size without a significant drop in accuracy. We started with a network (named CNN 2) specified in Fig. 5 and then decreased the number of convolutional blocks and dense layers as given in Tab. II. Each convolutional block had 16 filters of size  $5 \times 5$ , all images are downscaled to  $224 \times 224$ . As given in Tab. III the experiments showed that the information from the context could help the classification accuracy (or there is a strong correlation in the presence of weeds between the lines and

between the neighboring stems in the ROI area). While we can see a decreasing trend in accuracy from CNN 2 to CNN 5, the reduction of number of parameters is not significant (see Tab. II). Thus we made further variants of CNNs: reduced the number of neurons in dense layers and reduced the number of convolutions. In this process we generated 7 models, namely CNN 3.2, 3.3, 3.4, 3.5, 5.2.1, 5.2.2, and 5.3. The number of parameters and accuracy of these networks are visible in Tab. IV, Fig. 6, and Fig. 7. It is clear to see that there is a significant drop in accuracy for CNN 3.5, and halving the number of neurons in the dense layers of CNN 5.2.1 was not a good idea. Many of these simplified networks produces rather good results, the question is their computational power needed.

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 224, 224, 16)	1216
conv2d_1 (Conv2D)	(None, 224, 224, 16)	6416
conv2d_2 (Conv2D)	(None, 224, 224, 16)	6416
conv2d_3 (Conv2D)	(None, 224, 224, 16)	6416
max_pooling2d (MaxPooling2D)	(None, 112, 112, 16)	0
flatten (Flatten)	(None, 200704)	0
dense (Dense)	(None, 16)	3211280
dense_1 (Dense)	(None, 16)	272
dense_2 (Dense)	(None, 2)	34

-----  
 Total params: 3,232,050  
 Trainable params: 3,232,050  
 Non-trainable params: 0  
 -----

Fig. 5: The structure of the CNN 2 model.

TABLE II: Different base CNNs included in our study. Each convolutional layer has 16 filters.

Models	# parameters	# dense layers	# conv. layers
EffNetB0	4,779,045	3	
CNN 2	3,232,050	3	4
CNN 3	3,219,218	3	2
CNN 4	3,218,946	2	2
CNN 5	3,212,530	2	1

TABLE III: Accuracy of initial networks on images with/without masking.

	Whole images		Masked images		Test acc. diff.
	Val. acc.	Test acc.	Val. acc.	Test acc.	
EffNetB0	0,9977	1	0,9977	1	0
CNN 2	0,9909	0,9909	0,9658	0,9863	-0,0046
CNN 3	0,9932	0,9954	0,9408	0,968	-0,0274
CNN 4	0,9954	0,9863	0,9112	0,9452	-0,0411
CNN 5	0,9863	0,9954	0,7306	0,9315	-0,0639

TABLE IV: The main parameters and accuracy values of CNN 3 and CNN 5 variants.

	filters by layer	neurons in dense layers	params	tr.acc.	val.acc.	test.acc.
CNN3	16	16	3.219M	0.9741	0.9932	0.9954
CNN3.2	16	8	1,613M	0.9542	0.9818	1.0000
CNN3.3	8	8	805,130	0.9367	0.9658	0.9909
CNN3.4	4	4	201,446	0.8993	0.9431	0.9772
CNN3.5	2	2	50,444	0.5130	0.5239	0.2694
CNN 5	16	16	3.122M	0.9554	0.9863	0.9954
CNN 5.2.1	16	8	1,607M	0.5130	0.5239	0.2694
CNN 5.2.2	8	16	1,606M	0.9470	0.9863	1.0000
CNN 5.3	4	16	803,170	0.8981	0.9431	0.9817

## VI. PERFORMANCE ON EMBEDDED SYSTEMS

All in the previous experiments we used cloud services with massive GPU support which is not very typical in field applications often far from high-bandwidth networks. Luckily there are different embedded system platforms for application developers which could be operated in cultivators.

### A. Experiments on the Jetson AGX Xavier Development Platform

The NVIDIA Jetson AGX Xavier Series is an industrial platform for massively parallel computations reaching up to 32 TOPS. We run our tests on a 512 cores Volta architecture with 64 Tensor cores. As given in Tab. V, all models could run at high speed.

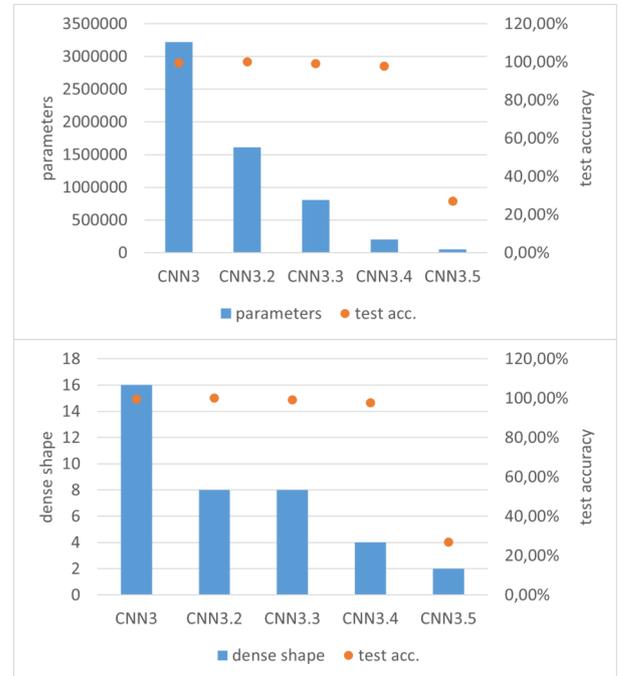


Fig. 6: Top: Accuracy and number of parameters of the CNN models created from model CNN 3. Bottom: Accuracy and shape of the same models.

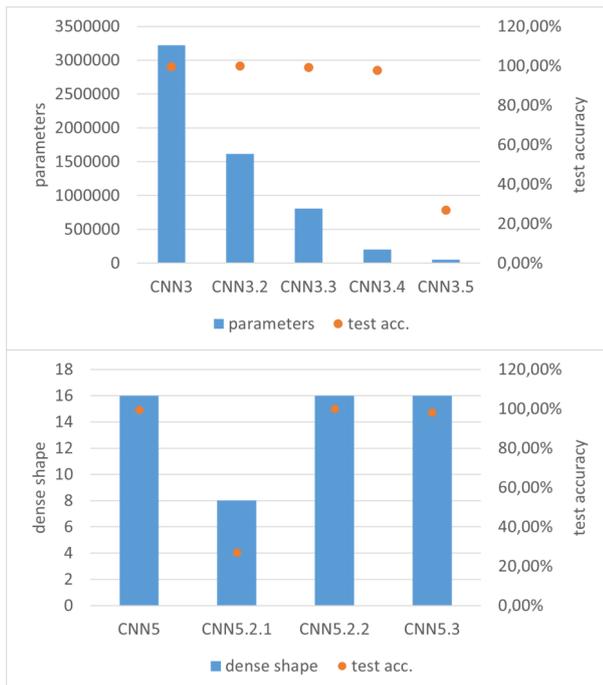


Fig. 7: Top: Accuracy and number of parameters of models created from model CNN 5. Bottom: Accuracy and shape of the same models.

TABLE V: The running performance of different CNN models on the Jetson AGX Xavier platform.

CNN Model	2	3.2	3.3	3.4	3.5	4	5	5.2	5.2.2	5.3
FPS	15.63	28.57	37.04	45.45	33.33	25.0	41.67	47.67	45.45	50

### B. Experiments on the STM32 Platform

STMicroelectronics produces different boards built on ARM cores which can be possible platforms for on-field weed recognition. It is possible to test different DNN models in a cloud service of STMicroelectronics at <https://stm32ai-cs.st.com/home>. Uploaded models can be optimized for speed, for memory usage, or for both. We have chosen the third option to test four of the previous CNN models on two platforms. According to Tab. VI acceptable FPS could be achieved with model CNN 5.3 only on the STM32H735G-DK platform.

TABLE VI: The running performance (FPS) of different CNN models on two specific STM32 boards.

CNN Model:	2	3.2	5.2	5.3
STM32H735G-DK	0.065	0.173	2.15	4.26
STM32F469I-DISCO	0.012	0.032	0.29	0.57

## VII. CONCLUSION

We have outlined a hybrid weed control approach where hoeing is combined with spraying, ensuring that the amount of applied chemicals is very low. We found that the localization of

lines can be achieved by Mask-R CNN segmentation of corns, while the recognition of weeds can be done with relatively small size CNNs. Considering the typical speed of cultivators both the NVidia AGX Xavier platform and the STM32H735G-DK board of STMicroelectronics are applicable for recognition. The detection of maize lines with Mask-R CNN on the Xavier platform is for future work as well as capturing new field images under different weather conditions and growth phases of corns. Since corn is the most produced cereal worldwide, our study shows that environmental friendly hybrid approaches can significantly contribute to reaching short term aims of agriculture.

## REFERENCES

- [1] Constanze Fetting, "The European Green Deal," ESDN Report, December 2020, ESDN Office, Vienna.
- [2] Aichen Wang, Wen Zhang, Xinhua Wei, "A review on weed detection using ground-based machine vision and image processing techniques," *Computers and Electronics in Agriculture*, vol. 158, 2019, pp. 226–240.
- [3] Food and Agriculture Organization of the United Nations, <https://www.fao.org/faostat>
- [4] Bakhshipour, A., Jafari, A., Nassiri, S.M., Zare, D., "Weed segmentation using texture features extracted from wavelet sub-images," *Biosyst. Eng.*, vol. 157, 2017, pp. 1–12.
- [5] Kumar, D.A., Prema, P., "A novel wrapping curvelet transformation based angular texture pattern (WCTATP) extraction method for weed identification," *ICTACT J. on Image Video Process.*, vol. 6, no. 3, 2016
- [6] García-Santillán, I., Guerrero, J.M., Montalvo, M., Pajares, G., "Curved and straight crop row detection by accumulation of green pixels from images in maize fields," *Precis. Agric.*, vol. 19, 2018, pp. 18–41.
- [7] Midtiby, H.S., Åstrand, B., Jørgensen, O., Jørgensen, R.N., "Upper limit for context-based crop classification in robotic weeding applications," *Biosyst. Eng.* vol. 146, 2016, pp. 183–192.
- [8] Xu, K., Li, H., Cao, W., Zhu, Y., Chen, R., and Ni, J., "Recognition of weeds in wheat fields based on the fusion of RGB images and depth images," *IEEE Access*, vol. 8, 2020, pp. 110362–110370.
- [9] Tang, J., Zhang, Z., Wang, D., Xin, J., He, L., "Research on weeds identification based on K-means feature learning," *Soft Comput.* vol. 22, 2018, pp. 7649–7658.
- [10] Hall, D., Dayoub, F., Kulk, J., McCool, C., "Towards unsupervised weed scouting for agricultural robotics," in *Robotics and Automation (ICRA)*, 2017 IEEE International Conference On. IEEE, pp. 5223–5230.
- [11] Szegedy, C., Liu W., Jia Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A., "Going Deeper with Convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.
- [12] Zhang, R., Wang, C., Hu, X., Liu, Y., and Chen, S., "Weed location and recognition based on UAV imaging and deep learning," *International Journal of Precision Agricultural Aviation*, 2020, vol. 3, no. 1, pp. 23–29.
- [13] Garibaldi-Márquez, F., Flores, G., Mercado-Ravell, D. A., Ramírez-Pedraza, A., and Valentín-Coronado, L. M., "Weed Classification from Natural Corn Field-Multi-Plant Images Based on Shallow and Deep Learning," *Sensors*, vol. 22, no. 8, 3021.
- [14] Simonyan, K.; Zisserman, A., "Very deep convolutional networks for large-scale image recognition," In *Proceedings of the 3rd International Conference on Learning Representations*, San Diego, USA, 7–9 May 2015.
- [15] Chollet, F. "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, USA, 21–26 July 2017, pp. 1800–1807.
- [16] Venkataraju, A., Arumugam, D., Stepan, C., Kiran, R., and Peters, T., "A Review of Machine Learning Techniques for Identifying Weeds in Corn," *Smart Agricultural Technology*, 2022, 100102.
- [17] He, K., Gkioxari, G., Dollár, P., and Girshick, R., "Mask R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2961–2969.
- [18] Tan, M., and Le, Q., "EfficientNet: Rethinking model scaling for convolutional neural networks," International Conference on Machine Learning, PMLR, 2019, pp. 6105–6114.