 DOI: 10.3217/978-3-85125-954-4-13

Crop row detection utilizing spatial CNN modules

Peter Riegler-Nurscher
Josephinum Research
Wieselburg, Austria
p.riegler-nurscher@josephinum.at

Leopold Rupp
CFS Cross Farm Solution GmbH
Stoitzendorf, Austria
lr@cfsolution.at

Abstract

Mechanical weed control is becoming increasingly important over conventional methods, not least because of environmental challenges. Precise guidance of the hoeing machine along the crop rows is necessary to be able to work efficiently. In this work, the use of deep learning methods for crop row detection is presented and evaluated on a custom data set. Recent advances in the task of vision based lane detection, like Spatial CNN (SCNN) and Recurrent Feature-Shift Aggregator (RESA), can potentially be applied to crop row detection as well. These methods are expected to improve the detection of the crop rows, especially in the case of strong weed growth and challenging environmental conditions, compared to the state of the art.

1. Introduction

There is a steadily increasing demand on the food market for food produced according to organic farming standards. Likewise, the proportion of organically farmed agricultural areas or organically managed arable farms in Europe is growing continuously. The change from chemical to mechanical weed control can only remain economical with a high degree of automation. Row hoeing equipment for weed removal often uses duck-foot shares that must be guided precisely along the row to prevent crop damage. State of the art camera systems used for row guidance have limitations, due to the robustness of conventional row detection algorithms especially with strong weed cover [6].

1.1. Related Work

Plant row detection in robotics as well as in marketed row guidance systems is based on conventional methods like line detection and color thresholds [4]. However, first approaches for convolutional neural network (CNN) based row recognition have already been presented [2]. Recent advances in the task of vision based lane detection could potentially be applied to crop row detection as well. These methods can be categorized into segmentation-based, point-based and curve-based lane detection methods [3]. Point-based methods directly output points whereas curve-based methods output curve parameters. Accordingly, other loss functions are used during training for point- and curve-based methods compared to segmentation based methods. Point- and curve-based methods are not discussed in detail in this paper and are also not included in the evaluation. Segmentation-based methods, like Spatial CNNs (SCNN) [7] and Recurrent Feature-Shift Aggregator (RESA) [9], output segmentation masks. A threshold is applied to the output to get sample discrete points on the lines. SCNN uses a spatial CNN module to model spatial relationships more efficiently than MRF or CRFs. The module is integrated after the top hidden layer. It preserves the continuity of long, thin structures over discontinuities. The RESA proposed in [9] utilizes spatial information by shifting sliced feature map. RESA is more computationally efficient than SCNN and also introduces an up-sampling decoder, the so called Bilateral Up-Sampling Decoder (BUSD). It is composed of two branches, a coarse grained branch and a fine detailed branch.

Methods for lane detection have not been utilized for crop row detection before. In this work, a new approach for crop row detection based on CNNs is presented. Additionally, the SCNN and RESA methods with different backbone CNNs are investigated for this task with a custom data set.

2. Method

Twelve different architectures for crop row detection were tested. As backbones ResNet [5] architectures of different sizes (ResNet18, ResNet34 and ResNet101), as well as a VGG16 [8] architecture are used. The up-sampling is done by a DeepLab [1] architecture except for the RESA method where the Bilateral Up-Sampling Decoder is used. The models take an input size of 800×288 pixel and outputs the segmentation mask in the same resolution. For augmentation, the training images were randomly flipped, rotated and a color jitter as well as random lighting were added. The backbone CNNs were pretrained on ImageNet. The implementation is built upon the framework introduced in [3]. All models were trained 120 epochs on the data set introduced in the following Section 3. At 120 epochs, convergence was observed for all variants.

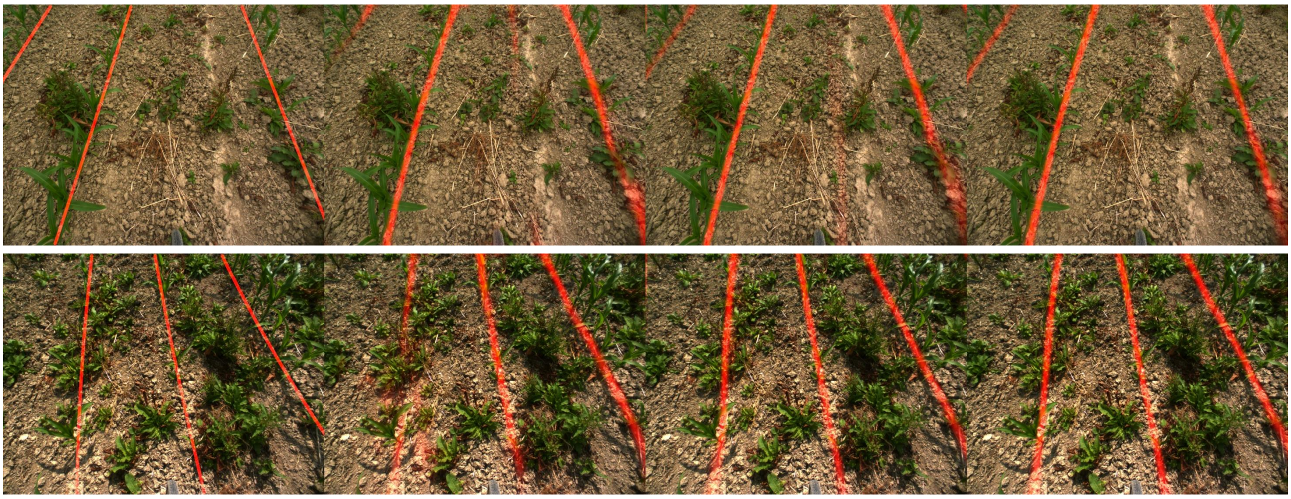

Figure 1. Segmentation result shown in red of different models (side by side) for two example test images (among one another).

3. Evaluation

For training and evaluation of the different methods, a custom data set was created. The data set consists of 3870 images of maize rows captured in the seasons 2021 and 2022 under various lighting conditions. The RGB images have a resolution of 1600×1200 pixels. The images are labelled with our custom labelling tool. A row can be defined by clicking a minimum of 2 points within a row. The row is afterwards interpolated by a polynomial of degree 2 from which regularly sampled points are stored. A segmentation mask is automatically generated by drawing curves with width of 16 pixel. All architectures have an input size of 800×288 pixel, therefore all images are resized to this resolution. The data set is split into 3475 images for training and 395 test images.

3.1. Results

The segmentation accuracy of the different methods on the test images is presented in Table 1. Figure 1 shows two examples of the test set with the segmentation results for ResNet18, ResNet101 and VGG16 with SCNN. The numbers presented, as well as the sample images, show an advantage in the use of SCNNs for detecting crop rows. The superiority of RESA over SCNN on lane detection data sets [3] could not be achieved in crop rows. Although, improvements of RESA over the baseline model (just Backbones with Deeplab) are visible. Likewise, it is recognizable that larger models, like ResNet101, achieve better detection rates without the use of spatial modules.

It might be assumed that the model is implicitly distinguishing crops from weeds based on test images with strong weed cover. However, this needs to be investigated in more detail.

	ResNet18	ResNet34	ResNet101	VGG16
		Baseline		
Accuracy	66.83	67.63	**67.90**	66.98
IoU	44.07	45.48	**44.82**	44.53
		SCNN		
Accuracy	68.15	69.18	68.58	**70.33**
IoU	44.19	44.93	44.83	**45.59**
		RESA		
Accuracy	66.51	**68.06**	56.93	N/A
IoU	43.58	**44.30**	40.66	

Table 1. Accuracy and Intersection over Union (IoU) of the crop row segmentation based on 395 test images.

4. Conclusion and Outlook

The work demonstrates the ability of CNNs for semantic segmentation to detect crop rows. Especially the SCNN but also the RESA method could improve the detection compared to the baseline methods. When selecting a method for steering a hoeing machine, however, the computational load should also be taken into account where SCNN has its drawbacks. Currently, we are working on integrating the models into a machine to steer along the rows. This allows for end to end evaluation of the system and an assessment of acceptable model errors. Future work could also focus on point- and curve-based methods.

Acknowledgments

This research was funded by FFG (Austrian Research Promotion Agency) under grant 880610 (DeepRow).

References

[1] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. In *Proceedings of the 3rd International Conference on Learning Representations, ICLR*, 2015.

[2] Rashed Doha, Mohammad Al Hasan, Sohel Anwar, and Veera Rajendran. Deep learning based crop row detection with online domain adaptation. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '21, page 2773–2781, 2021.

[3] Zhengyang Feng, Shaohua Guo, Xin Tan, Ke Xu, Min Wang, and Lizhuang Ma. Rethinking efficient lane detection via curve modeling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, pages 17062–17070, 2022.

[4] Chuangxin He, Qingtai Chen, Zhonghua Miao, Nan Li, and Teng Sun. Extracting the navigation path of an agricultural plant protection robot based on machine vision. In *Proceedings of the 40th Chinese Control Conference, CCC*, pages 3576–3581, 2021.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, pages 770–778, 2016.

[6] Oskar Kress, Sabine Staub, and Simon Brell. Abschlussbericht Forschungsprojekt "Beikrautregulierung in Ökobetrieben mit Gemüsekulturen unter besonderer Betrachtung von moderner RTK-Steuerungs-, Ultraschall- und Kameratechnik inkl. Arbeitswirtschaft und Kosten". Bayerische Landesanstalt für Weinbau und Gartenbau (LWG), 2019.

[7] Xingang Pan, Jianping Shi, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Spatial as deep: Spatial cnn for traffic scene understanding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, pages 7276–7283, 2018.

[8] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *Proceedings of the 3rd International Conference on Learning Representations, ICLR*, 2015.

[9] Tu Zheng, Hao Fang, Yi Zhang, Wenjian Tang, Zheng Yang, Haifeng Liu, and Deng Cai. Resa: Recurrent feature-shift aggregator for lane detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 3547–3554, 2021.