Using Fully Convolutional Networks for Rumex Obtusifolius segmentation, a preliminary Report

Schori Damian[1], Anken Thomas[2], Šeatović Dejan[1]
[1]University of Applied Science Rapperswil (ILT Institute for Lab Automation and Mechatronics), Switzerland
[2]Agroscope, Tänikon 1, CH-8356 Ettenhausen, Switzerland

damian.schori@hsr.ch

***Abstract* - Image segmentation of specific plants is an important task in precision farming. Several influences such as changing light, varying arrangement of leaves and similarly looking plants are challenging. We present a solution for segmenting individual Rumex obtusifolius plants out of complicated natural scenes in grassland from 2D images. We are making use of a fully convolutional deep neural network (FCN) trained with hand labeled images. The proposed segmentation scheme is validated with images taken under outdoor conditions.**

The overall masks segmentation rate is 84.8% measured by the dice coefficient. Approximately half of the experiments show segmentation rates of individual plants higher than 88%. The developed solution is therefore a robust method to segment Rumex obtusifolius plants under real-world conditions in short time.

***Keywords* - Deep Learning, Image Segmentation, Computer Vision, Rumex obtusifolius**

I. Introduction

The detection of specific plants and their corresponding leaves is one of the essential tasks for the automation of plant specific treatments. While the research on organic treatment of Rumex obtusifolius has constantly been improved and the detection of R. obtusifolius stalled.

Two solution paths have been researched. The major path is based on a variety of image analysis approaches in 2D including Fourier Analysis and Sliding Window Classifier [1, 2, 3, 4]. The second solution approach focuses on the 3D point clouds analysis and had promising results [5, 6, 7]. A comprehensive summary of current state of the art detection solutions is explained by Binch and Fox [8].

Due to recent advances in Deep Learning, exploration of new processing procedures became far less time consuming. In this paper we present preliminary and promising results towards the solution of real-time plant segmentation and treatment in grassland using deep learning. The results originate from a student project, WeedEraser: An autonomous ground vehicle for weed detection and treatment in cooperation with Agroscope, Swiss Federal Office for Agriculture.

II. Problem description

The segmentation of individual plants in their natural environment is an extremely challenging task due to plants showing significantly varying poses, sizes and complex shapes in natural environmental conditions.

R. obtusifolius differs clearly from most weed in grassland by its shape and size. However, there are plants like Plantago lanceolata or Taraxacum officinale, which show significant similarities in shape and size that influence the detection process. Furthermore the habitus of R. obtusifolius varies extremely in function of it's natural environment and appears in different sizes and constellations in a field. Additional factors such as changing light and field properties influence the detection.

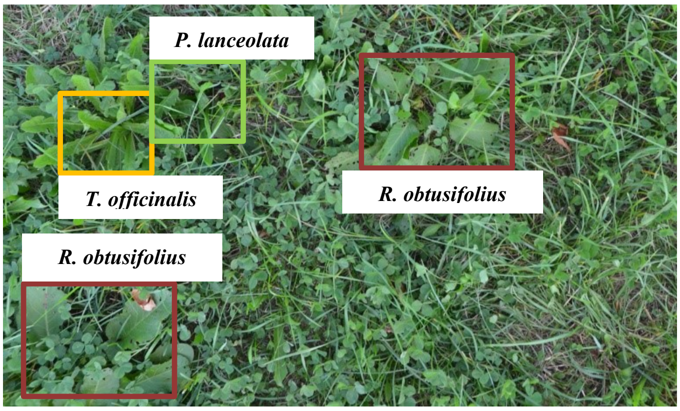

Fig. 1. *Detail of a training image containing similar plants in size and shape to R. obtusifolius.*

Therefore a major goal is not only to identify the leaves but exclude similar species reliably and minimize the treatment time. The treatment of the plant is purely organic: The hot water jet (80 °C, 120– 130 bar) penetrates the ground up to 15cm and heats the plant root area above 60 °C. Detailed treatment description can be found in [9].

First robotic systems for R. obtusifolius detection and treatment were realized more than ten years ago and the research lasts till today (2019) [1, 2, 3, 10, 11]. A major issue of these systems are low detection rates due to suboptimal segmentation/ classification algorithms.

978-1-7281-2181-9/19/$31.00 ©2019 IEEE

III. Model Architecture

A. Training Data

Field images gained on several pastures in Switzerland containing R.obtusifolius were provided as 1024x1536 pixel images. For the leaf segmentation, a hand-labeled binary mask was created for each image (Fig. 2). The training and validation dataset consist of 507 images.

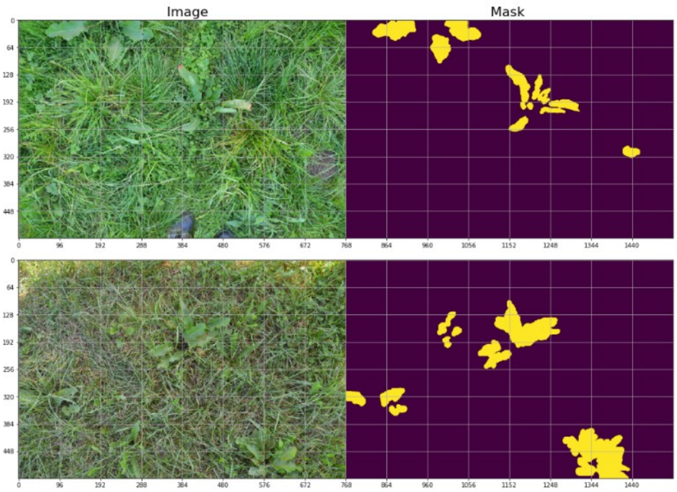

***Fig.* 2**. *Samples of the dataset: field images and ground truth masks.*

The images were taken by hand with an RGB- Camera from about one meter above ground. This resulted in a recording area of approx. 1.2 x 1.8m. For training we resized the images to 512x768 pixels. Higher resolutions didn't improve accuracy and resulted in longer training times.

On each input image the pixel intensity is normalized with a mean of 0.5 and a standard deviation of 0.1 over all color channels.

B. Model Architecture

The proposed plant segmentation scheme is mainly making use of an adapted version of U-Net (Fig. 3), a Fully Convolutional Neural Network (FCNs) proposed in [12]. FCNs were introduced [13, 12, 14] in the literature as a natural extension of CNNs to tackle per pixel prediction problems such as semantic image segmentation. FCNs add up-sampling layers to standard CNNs to recover the spatial resolution of the input at the output layer. Consequently, FCNs can process images of arbitrary size. In order to compensate the resolution loss induced by pooling layers, FCNs introduce skip connections between their down-sampling (encoder) and up-sampling (decoder) paths. Skip connections help the up-sampling path recover fine-grained information from the down-sampling layers.

The proposed network features an encoder path with six resolution levels, which reduce the spatial dimensions of the input image from 512x768 to 16x24 pixels at the lowest resolution and encode the relevant context and information into 1024-dimensional feature vectors. The decoding path is used to enable precise localization of the leaves using upsampling layers.

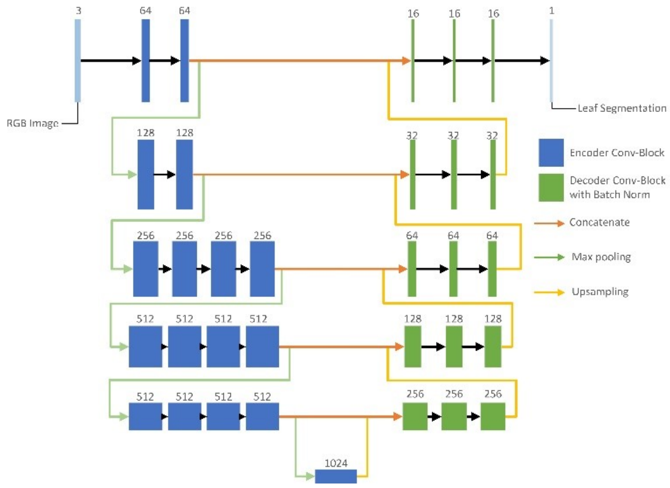

***Fig.* 3**. *Proposed Network Architecture based on Unet [12]*

C. Training

For better and faster convergence, a VGG19 Network [15] pretrained on ImageNet [16] was used as the encoder part. We trained the network in the Keras-Framework [17] with a TensorFlow [18] backend until convergence (50 epochs) using the ADAM solver [19] with the first momentum set to 0.9 and the second to 0.99 and a learning rate decay of 0.1 after every 20 epochs.

The loss function is defined as a combination of binary cross entropy and the dice coefficient. It is calculated by a pixel-wise sigmoid activation over the output-feature map. The loss-function is there for defined as:

$$H = H_{BCE} + H_{Dice}$$

where:

$$H_{BCE} = -\sum_{i=0}^{M} (y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i))$$

$$H_{Dice} = -\frac{2 \cdot \sum_{i=0}^{M} (\hat{y}_i \cdot y_i)}{\sum_{i=0}^{M} (\hat{y}_i + y_i)}$$

- Let $i = \{0 \,..\, M\}$ denote the indices of the flatted vectors of y and $\hat{y}$.

D. Data Augmentation

The network has a large number of trainable parameters (about 36 million) compared to the amount of available training images. To teach the network the desired invariance and robustness properties we made use of real time data augmentation [20] during training time.

In our case of field-images we primarily need shift and rotation invariance as well as robustness to light, size and bluring variations. The number of augmentation operations applied is chosen randomly from zero to four operations per image.

A single Nvidia GTX 1070 GPU was used for the training of the described model. The full training and prediction cycle took about four hours.

E. Evaluation metrics

The algorithm was evaluated by computing image wise dice coefficient (DCs), pixel-based true positive (TPs) and false positive (FPs) rates on the "good" segmentations with DC > t and object-based false negative (FNs), measuring the rate of "bad" segmentations (DCs ≤ t) where t = 0.7 is the DCs threshold.

The metrics are therefore defined as follows:

- $DCs = \frac{2 \cdot \sum_{i=0}^{M}(\hat{y}_i \cdot y_i)}{\sum_{i=0}^{M}(\hat{y}_i + y_i)}$
- $TPs = \frac{\sum_{i=0}^{M}(\hat{y}_i \cdot y_i)}{M}$
- $FPs = \frac{\sum_{i=0}^{M}[\hat{y}_i = 1 \wedge y_i = 0]}{M}$
- $FNs = [DC \leq t]$
- Let y denote a vector of the flatted ground-truth mask
- Let $\hat{y}$ denote a vector of the flatted prediction mask with pixel values which have probability higher than 0.7
- Let $i = \{0 \mathinner{..} M\}$ denote the indices of the flatted vectors of y and $\hat{y}$

IV. Results

The 507 images were once split to 20% training and 80% validation images and once to 80% training and 20% validation images. Pixels with detected Rumex were classified into TP (true positive = correct detection where there is a Rumex in the pixel) and FP (detection of a Rumex where is none in that pixel) and FN (missed out Rumex, i.e. DC < 0.7). Using 20% of the images for training and 80% for validation already resulted in a dice coefficient (DC) value of 80%. Increasing the training set decreased the number of missed Rumex areas (FN) drastically.

	DC	*TP*	*FP*	*FN*
20% training / 80% validation	**0.80** ±0.18	**0.98** ±0.017	**0.007** ±0.009	**0.15** ±0.35
80% training / 20% validation	**0.84** ±0.12	**0.99** ±0.007	**0.006** ±0.007	**0.05** ±0.23

Table 1. *Classification of the results (average over all images and standard deviation)*

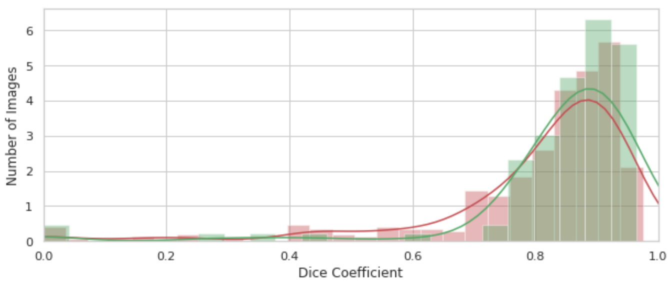

Fig. 4. *Distribution of the calculated Dice Coefficient of the Validation Images.* ***Red****: 20% training and 80% validation split.* ***Green****: 80% training and 20% validation split.*

Fig. 4 shows the distribution of the dice coefficient over all validation images for both dataset splits. The 20/80 split shows slightly worse DC results and significantly more false negatives. Most mistakes are made on images containing very little area of R.obtusifolius.

Fig. 5. *Sample prediction on Validation Data (80% training / 20% validation).* ***(Left)*** *Field Images as Input.* ***(Right)*** *Predicted segmentation result (yellow mask) with framed ground truth (cyan border)*

V. Conclusion

We can show that our deep learning approach is able to segment R.obtusifolius plants with good accuracy from grassland and other similar looking plants. In earlier works [8], which are mostly based on Fourier Analysis, appealing results were achieved, but these depend on the fact that roughly the same light situations always prevail.

However, our method is due to variations in the training set and additionally added light augmentation robust to changing light situations such as sunlight or shadows (Fig. 5). Also, the method can deal with partly covered leaves by other plants and softly blurred images.

But considering the results, there is still optimization potential which will be investigated in further work. This also includes locating the exact root position of the plant as well as training and evaluating the network on a larger data set. Moreover, the speed of the algorithm must be considered, as the plants must be detected in real time during the meadow crossing.

VI. Acknowledgments

The authors would like to thank the Data Science community for many valuable insights and educational help in the field of machine/deep learning.

Furthermore we want to express our appreciation to companies that support our work: Omron Switzerland, Beckhoff Switzerland and NVIDIA Switzerland. Without their support our project WeedEraser would still be a vision.

REFERENCES

[1] F. Van Evert, G. Polder, G. Van der Heijden, C. Kempenaar and L. Plotz, “Real-time vision-based detection of Rumex obtusifolius,” p. 11, 2008.

[2] F. J. S. G. P. M. V. H.-J. v. D. A. L. e. a. Van Evert, „A robot to detect and control broad-leaved dock (Rumex obtusifolius L.) in grassland.,“ *Journal of Field Robotics,* Nr. 28, pp. 264-277.

[3] F. G. P. G. V. d. H. C. K. a. L. L. Van Evert, „Real-time, vision-based detection of Rumex obtusifolius L. in grassland.,“ *Weed Research,* Nr. 49, pp. 164-174, 2009.

[4] T. Kounalakis, G. A. Triantafyllidis und L. Nalpantidis, „Deep Learning-based Visual Recognition of Rumex for Robotic Precision Farming,“ p. 32, 2019.

[5] D. Šeatović, „A Segmentation Approach in Novel Real Time 3D Plant Recognition System,“ in *Computer Vision Systems*, Bd. 5008, A. Gasteratos, M. Vincze und J. Tsotsos, Hrsg., Springer Berlin / Heidelberg, 2008, pp. 363-372.

[6] D. Šeatović, H. Kutterer und T. Anken, „Automatic weed detection and treatment in grasslands,“ in *ELMAR, 2010 PROCEEDINGS*, 2010.

[7] D. Šeatović, „Methods for Real Time Plant Detection in 3-D Point Clouds,“ 2013.

[8] A. Binch und C. Fox, „Controlled comparison of machine vision algorithms for Rumex and Urtica detection in grassland,“ 2017.

[9] R. Latsch und J. Sauter, „Optimisation of hot-water application technology for the control of broad-leaved dock (Rumex obtusifolius),“ *Journal of Agricultural Engineering,* Bd. 45, p. 137, 12 2014.

[10] D. Šeatović und R. Grüninger, „Smart Weeder: Novel Approach In 3D Object Recognition, Localization And Treatement Of Broad Dock In Its Natural Enviroment,“ 2007.

[11] F. A. L. A. D. J. E. G. G. V. d. H. L. L. Van Evert, „Sietse III: Or: Robot recognizes Rumex.,“ in *Proceedings of the 4th Field Robot Event 2006*, Stuttgart/Hohenheim, 2009.

[12] O. Ronneberger , P. Fischer und T. Brox, „U-Net: Convolutional Networks for Biomedical Image Segmentation,“ p. 8, 2015.

[13] J. Long, E. Shelhamer und T. Darell, „Fully Convolutional Networks for Semantic Segmentation,“ *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* p. 10, 2015.

[14] S. Jégou, M. Drozdal, D. Vazquez, A. Romero und Y. Bendio, „The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation,“ 2017.

[15] K. Simonyan und A. Zisserman, „Very Deep Convolutional Networks for Large-Scale Image Recognition,“ *arXiv,* p. 14, 2014.

[16] „ImageNet,“ May 2019. [Online]. Available: http://www.image-net.org/.

[17] „TensorFlow,“ May 2019. [Online]. Available: https://www.tensorflow.org/.

[18] F. Chollet und others, „Keras,“ May 2019. [Online]. Available: https://keras.io.

[19] D. P. Kingma und J. Lei Ba, „Adam: A Method For Stochastic Optimization,“ *ICLR,* 2015.

[20] „Imgaug,“ 04 2019. [Online]. Available: https://github.com/aleju/imgaug.

[21] K. Itakura und F. Hosoi, „Automatic Leaf Segmentation for Estimating Leaf Area and Leaf Inclination Angle in 3D Plant Images,“ p. 11, 2018.

[22] J. Redmon, S. Divvala, R. Girshick und A. Farhadi, „You Only Look Once: Unified, Real-Time Object Detection,“ p. 10, 2015.

[23] S. G. V. d. H. F. V. E. a. A. S. Hiremath, „The role of textures to improve the detection accuracy of Rumex obtusifolius in robotic systems.,“ Weed Research, 2012.

[24] F. J. S. G. P. M. V. H. V. D. E. L. e. a. Van Evert, „Robotic control of broad-leaved dock.,“ *Precision Agriculture,* pp. 725-732, 2009.