Advances in Animal Biosciences: Precision Agriculture (ECPA) 2017, (2017), **8:2**, pp 842–847 © The Animal Consortium 2017
doi:10.1017/S2040470017000206

advances in animal biosciences

RoboWeedSupport - Detection of weed locations in leaf occluded cereal crops using a fully convolutional neural network

M. Dyrmann[1†], R. N. Jørgensen[2] and H. S. Midtiby[1]

[1]*The Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark;* [2]*Department of Engineering – Signal Processing, Faculty of Science and Technology, Aarhus University, Finlandsgade 22, 8200 Aarhus N, Denmark*

This paper presents a method for automating weed detection in colour images despite heavy leaf occlusion. A fully convolutional neural network is used to detect the weeds. The network is trained and validated on a total of more than 17,000 annotations of weeds in images from winter wheat fields, which have been collected using a camera mounted on an all-terrain vehicle. Hereby, the network is able to automatically detect single weed instances in cereal fields despite heavy leaf occlusion.

Keywords: Weed Detection, Integrated Weed Management, Deep Learning, Convolutional Neural Network

Introduction

In order to conduct site specific weed management, detailed information about the weed infestation regarding position composition within agricultural fields is required. The composition of weed species can be inferred from images acquired in the field.

When conducting chemical treatment of weeds, it is necessary to know the composition of the weed species in order to dose the herbicides optimally. When conducting mechanical weed control, it is necessary to know the location of the crop plants or the weeds in order to protect the crop plants while being able to remove the weeds effectively with a minimal use of energy.

In Northern Europe[1], cereals[2] are the primary crops, accounting for 52% of the total crop area (Eurostat, 2016). Therefore, being able to detect weeds in cereal fields will be valuable for precision weed control. However, most current methods for automated detection of weeds are designed for high value crops or crops with a large row distance, which allow for inspection between the crop rows, contrary to cereal, which typically is sown with a row distance of only 12 cm, which complicate between-row inspections. Many existing methods for weed detection and classification are based on images that are segmented using excess green or near-infrared recordings, where each object in the segmented images is assumed to be a single plant (Pérez *et al*. 2000; Giselsson 2014). This assumption may be sufficient in fields with a low degree of overlapping plants. However, for highly occluded fields, it is necessary to use methods that are capable of detecting weeds despite this occlusion. Such methods exist, including the methods by Andújar *et al*. (2012), who uses ultrasonic sound to estimate weed pressure, and Herrmann *et al*. (2012), Lottes *et al*. (2016) and Dyrmann *et al*. (2016) who make a pixel-wise segmentation of weeds and crops. The method developed by Herrmann *et al*. uses partial least square discriminant analysis for segmenting broad and narrows leaved weeds from crops and soil. The method by Lottes *et al*. is able to detect weeds in sugar beet fields by using a combination of global and local texture features, whereas the method by Dyrmann *et al*. is able to detect weeds in maize fields using a fully convolutional neural network. Another approach is the one by Laursen *et al*. (2016), which is able to quantify the area covered by dicots and monocots in images by analyzing orientations of leaf edge segments relative to one another.

Common for these methods is that they are not verified to work in cereal fields or that they are only able to quantify coverage of the crops and weeds, but not identify individual instances of the weeds. However, it is important to be able to detect either individual instances of weeds, individual instances of crops or the crop rows, if the weed control is to be carried out mechanically, or if the weed species is to be determined using weed classification methods, designed for single instance classification. The challenge is therefore to find a technique for weed detection that is able to handle weeds in cereal fields, in which overlapping plants often occur because of narrow row spacing. In this paper, we present a method for weed instance detection that can detect weeds in highly occluded cereal fields. The resulting method is based on DetectNet by Nvidia (Barker, J. *et al*., 2016), which again is based on GoogLeNet (Szegedy *et al*., 2014) that has been modified to a fully convolutional neural network, thereby producing images rather than labels.

† E-mail: mady@mmmi.sdu.dk

[1]Countries defined as Northern Europe by the United Nations Statistics Division: Denmark,Estonia, Finland, Iceland, Ireland, Latvia, Lithuania, Norway, Sweden and United Kingdom

[2]Barley, Durum wheat, Oats, Rye, Wheat and spelt, maslin, Triticale, and other cereals n.e.c.(buckwheat, millet, canary seed, etc.)

https://doi.org/10.1017/S2040470017000206 Published online by Cambridge University Press

Materials and Methods

In order to train the networks to be able to detect individual plants, a vast amount of training images are required. The collection of these images is carried out with a camera mounted on ATVs (All-Terrain Vehicles) as described by Laursen *et al.*, (2016a). This has resulted in a dataset consisting of 1,427 images from 10 fields, in which 18,541 weeds have been annotated manually. The annotation of these weeds is made using a web portal, as described in Rydahl *et al.* (2016). Here, the user can mark each weed instance in the image and denote the species of each of the weeds. In this study, the weed species are not used, but only the markings of the location of the weeds in the images, which are translated into a bounding box for each annotated plant in the images. The images are subsequently scaled down to 1224×1024 pixels. This size limit is because the Nvidia TitanX GPU used for training would otherwise run low on memory.

One of the 10 fields is selected for validation, and it is thus not included in the training of the model. By choosing an entire field for validation, we achieve a greater independence of the test and validation data, compared to a random selection of all the annotated images. The field that is selected for validation is the field in which the wheat is at the latest growth stage of the 10 fields. This field has a severe degree of occlusion compared to the others. By choosing the field in which there is greatest degree of occlusion, we risk reducing the system's performance, but in return, we make clear one of the big challenges with automatic detection of weeds, which has to be solved in order to carry out precision weed control in cereal fields. The validation dataset consists of 153 images, containing 876 annotated weeds. Image samples from the training and validation-set can be seen in Figure 1.

Network architecture

The convolutional neural network used for training is based on DetectNet by Nvidia (Barker, J. *et al.*, 2016). It consists of two parts: A fully convolutional neural network and a clustering function as illustrated in Figure 2. The fully convolutional neural network is trained by evaluating the errors in predicted bounding boxes and a predicted coverage map, which indicates detected locations of weeds in the image.

The fully convolutional neural network is based on the GoogLeNet architecture, from which the input layer and final pooling layer and inner product layer have been removed. Using an existing network allows one to use pre-trained models to initialize the network weights. Oquab *et al.* (2014) show that initializing convolutional neural networks with a model trained on a different data set causes the network to converge faster despite the fact that there are only few visual resemblances between the data set used for pretraining and the one used for training. GoogLeNet was designed for the ImageNet Challenge (Russakovsky *et al.* 2015). Therefore a model trained on the ImageNet data is used for initializing the weight in this study.

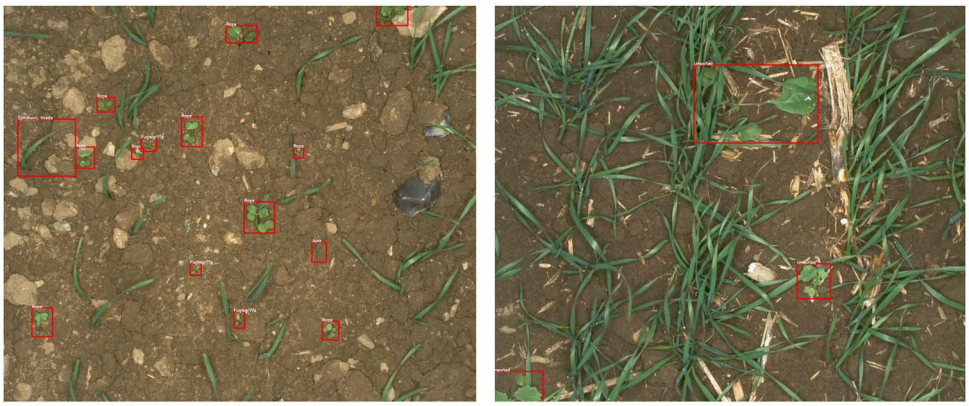


Figure 1 Hand-annotated samples from (left) the training data and (right) the validation data. Notice the amount of occlusion in the image to the right.

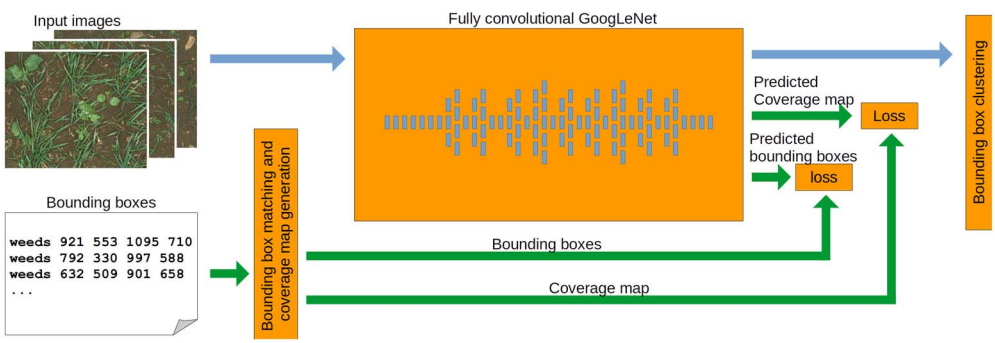


Figure 2 Network architecture. Input images and hand-made bounding boxed are fed into the network which produces a coverage map and bounding boxes. Green connections are only active during training. (after Barker, J. et al., (2016))

https://doi.org/10.1017/S2040470017000206 Published online by Cambridge University Press

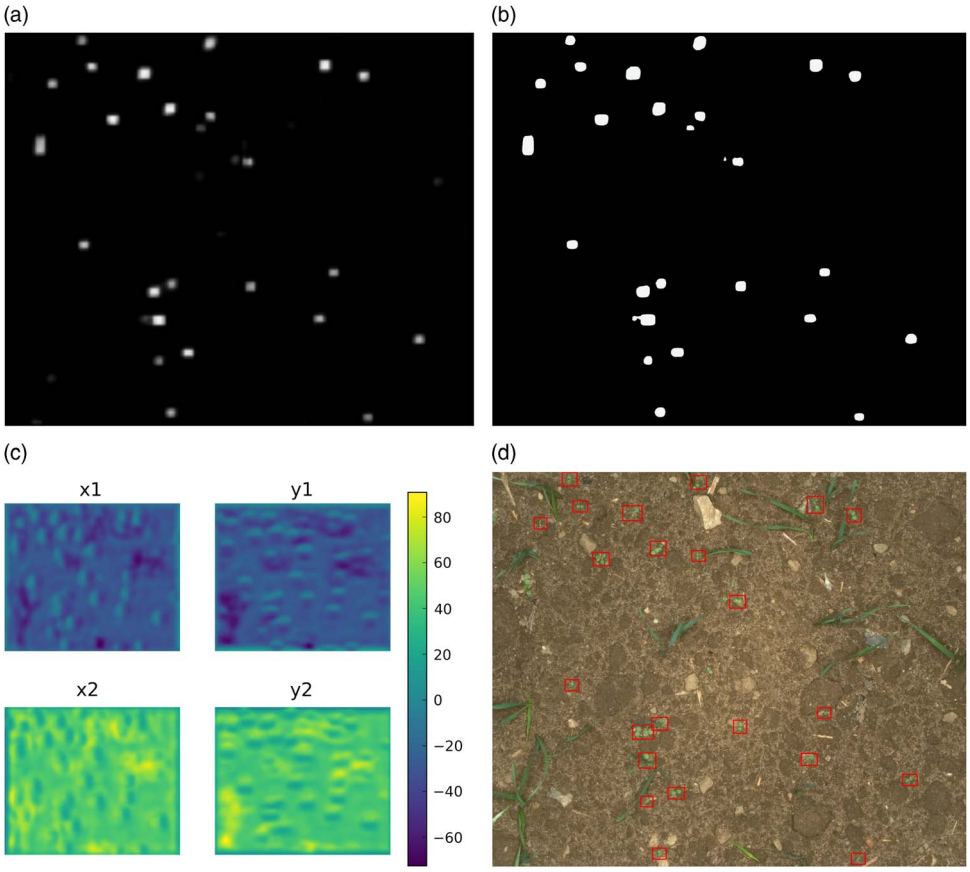


Figure 3 Procedure of the bounding box generation. (a) coverage map produced by network. (b) threshold of (a). (c) offsets of x1,y1,x2 and y2 relative to white pixels in (b). (d) final annotated image.

Network input and output

Images in which weeds are annotated with bounding boxes are required for training the network. These bounding boxes are used for generating a coverage map, which shows whether weeds are present at a given location in the image. This input coverage map for training is considered the "ground-truth" that we want the network output to match. The input coverage map is generated by rounding the corner coordinates of all bounding boxes to the nearest integer that is divisible by 16. Thereby each bounding box will fill a certain number of cells in a grid. Each cell in the grid are then assigned to a coverage value of 0 or 1, depending on whether a bounding box covers the grid cell. If the cell is covered, it is given a value of 1, otherwise 0. When testing the network, an image is the only required input.

As output, the fully convolutional neural network produces a new coverage map, in which each pixel corresponds to an area in the input image like for the input coverage map. This output coverage map shows how much the network believes that a plant is present at this location in the image. Along with the coverage map, the network outputs four bounding-box images. Each of these four bounding-box images contains pixel-values corresponding to the corner-coordinates of the bounding boxes.

Bounding box clustering

As described above, the network generates an output of five images; One coverage map and four bounding-box images. Here we will describe how these images are translated into bounding-boxes that mark the location of plants in the images.

Figure 3 shows an example of a generated coverage map as well as the four bounding-box images. The first step is to make a binary threshold of the coverage map. The specific threshold value is not of great importance, as it is being set before training the network, which means that the network will be able to adapt its internal biases for the given threshold value. However, after training the network, the threshold can be lowered, which will cause the network to accept uncertain detections, or increased, which will cause the network to keep only the most certain detections.

For each white pixel in the thresholded coverage map, a bounding box is created. Each bounding box is specified by two coordinates, (x1,y1) and (x2,y2), defining the two opposite corners of the bounding box. The values of these two coordinates are extracted from the four bounding box-images, shown in Figure 3(c). The pixel values at a given location in each of these four images are the offsets of respectively x1, y1, x2 and y4, relative to the location of a white pixel in the coverage map.

As each cluster in the coverage map consists of several pixels, a lot of bounding boxes are generated, for each cluster. Therefore, in order to only get one bounding box for each plant, the bounding boxes are grouped together if the relative difference between the sides of the rectangles is less

https://doi.org/10.1017/S2040470017000206 Published online by Cambridge University Press

than a given threshold. Setting this threshold too high results in multiple detections of single plants, whereas setting it too low causes overlapping plants to be detected as one.

Network losses

The fully convolutional neural network and the bounding box predictor are trained simultaneously, where only the output of the bounding box predictor is used as cost for training both the network and the clustering function.

The sum of two cost functions is used for optimizing the network weights during the training phase. These losses are a coverage loss and a bounding box loss. The coverage loss is the mean square error of the coverage map predicted by the fully convolutional neural network and the coverage map generated from the annotated data. Because of this loss, all weeds in the training images must be annotated, as a missing plant in the training data will cause the coverage loss to increase.

$$COV_loss = \frac{1}{2N}\sum_{i}^{i=N} \left| CM_{inp}(i) - CM_{pre}(i) \right|^2$$

The other loss is the bounding box loss, which is the mean absolute difference of corners of the predicted bounding boxes and the input bounding boxes.

$$BBOX_loss = \frac{1}{2N}\sum_{i}^{i=N} \left(\left| x_{inp}^{min} - x_{pre}^{min} \right| + \left| x_{inp}^{max} - x_{pre}^{max} \right| + \left| y_{inp}^{min} - y_{pre}^{min} \right| + \left| y_{inp}^{max} - y_{pre}^{max} \right| \right)$$

These losses are calculated for each of the training images, but in order to stabilize the weight update, the losses are averaged over eight training images.

Evaluation

In order to evaluate the system's ability to detect weeds, we need to set some matrix that indicates whether a weed is detected or not. To do so, we evaluate the Intersection Over Union (IOU), which evaluate the area a detected bounding box and an annotated bounding box have in common in proportion to the area that the annotation and detection cover together.

A plant is considered detected if the IOU is greater than ⅓, corresponding to a 50% overlap for two equal sized bounding-boxes, and if the center of the detection bounding box falls within the bounding box of the annotation and vice-versa. However, it is not enough to only consider the average IOU, since the IOU can be high while only few plants are detected. Therefore, we also measure the precision and recall.

The precision indicates the proportion of true positives relative to all detections. A high precision means that a large proportion of objects, identified as weeds by the algorithm, are also annotated as plants in the ground truth annotations. The recall, on the other hand, indicates the proportion of annotated plants that the algorithm has found automatically.

Results

A total of 876 plants were annotated in the 153 validation images. During the evaluation, 406 of these 876 annotated plants were detected automatically by the algorithm. This yielded a recall of 46.3% and a precision of 86.6%. The average IOU for accepted detection is 0.64. The mean size of all the hand-annotated bounding boxes is 12.064 pixels, ranging from 702 pixels for the smallest annotation up to 247.123 pixels for the largest annotation. However, the mean size of the automatically detected bounding boxes is 15.006 pixels, which indicates that the algorithm encounters more difficulties in detecting smaller plants than it does in detecting the larger ones. The mean size of the bounding boxes of the plants that were not detected, is, therefore, a bit lower, with an average size of 9.534 pixels. Samples of the test image where the algorithm fail to detect the plants and detect all the plants is shown in Figure 4.

Plants that have not been detected, according to the IOU criteria, can generally be divided into four groups: small plants, plants exposed to heavy occlusion, grasses, and plants where only part of the plant has been detected. Figure 5 shows crops of the validation images, which exemplifies each of these cases: (a) small plant, (b) plant exposed to heavy occlusion, (c) grasses, (d), plants, where only part of it has been detected. Testing the method was

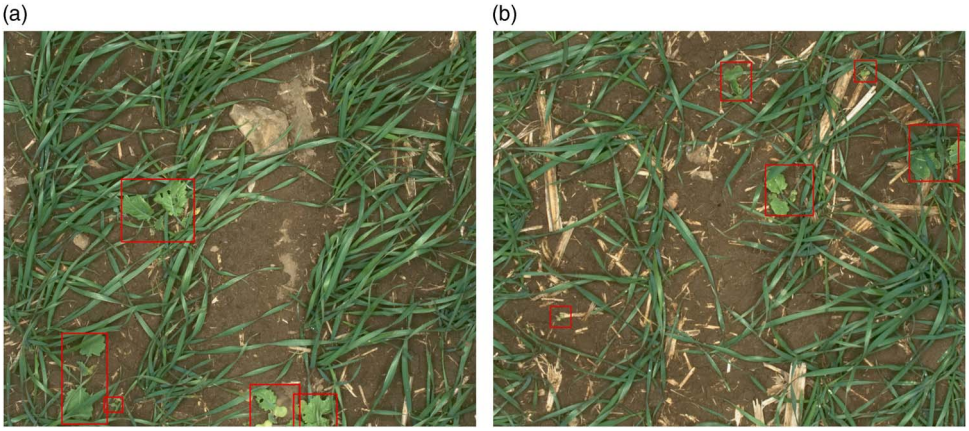


Figure 4 Samples from the validation dataset, where the algorithm (a) fails to detect the weeds and (b) detects all the weeds. The annotations in (a) is hand-made annotations in order to show the location of the missing plants. The annotations in (b) are made by the algorithm.

https://doi.org/10.1017/S2040470017000206 Published online by Cambridge University Press

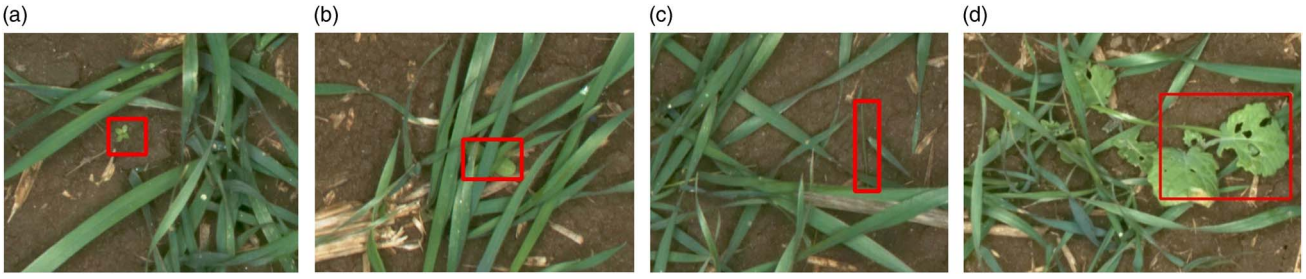


Figure 5 Plants that the algorithm fails to detect. (a) small plant, (b) plant exposed to heavy occlusion, (c) grasses, (d), plants, where only part of it has been detected. Markings in (a),(b) and (c) are hand-annotations, marking in (d) is made by the algorithm.

carried out using an Nvidia TitanX (Pascal) GPU, which was able to process 20 images per second.

Discussion

Weed instances in cereal fields have been hard to detect automatically, as cereals are sown with a low distance, causing crops to overlap with weeds. Furthermore, the narrow row distance makes it hard to use spatial information for detecting the weeds as is possible for crops sown with larger row spacing, such as maize. Cereal is, however, the main crop in northern Europe, so to be able to perform precision weed control in cereal fields is essential. Therefore, one of the greatest accomplishments in this study is the ability to detect weed instances in winter wheat fields. The field, which was selected for validation, was a field with a severe degree of occlusion, which is partly the reason why not all weeds were detected.

It is believed, that the performance can be increased, by training the network with more images that contain overlapping plants. When training the network, it is important, that all weeds are annotated in the training data, as a detection of a plant, that is not annotated, will count as an error, causing the network to learn invalid pattern in the data. In the training data, which was used in this study, most weeds are annotated, but not all. A further increase of the recall is therefore expected, if the training data was fully annotated.

Detecting weeds is, however, not enough to determine optimal herbicide dosages, if the weed control is to be carried out chemically. In that case the network will be the first step of a two-step operation for determining plant species and weed infestation of fields, where the other step is an algorithm optimized for classification of the species, such as the method described in Dyrmann *et al.*, (2016). When selecting herbicides, it is important that grasses are not missed by the algorithm, as grasses are a strong competitor to the crops. However, detecting grasses is a challenge, due to the many visual similarities grass has with wheat. Further development is therefore necessary to make the algorithm better at detecting grasses.

Conclusion

In this paper a method capable of detecting weed instances in highly occluded cereal fields has been presented. Current methods for weed detection have, so far, had trouble with detecting weed instances in cereal fields due to heavy leaf occlusion. We addressed this problem of detecting weed instances in cereal fields by using a fully convolutional neural network that is based on a modified version of the GoogLeNet architecture. The network has been trained on thousands of annotated images, which has enabled the network to detect single weed instances in cereal fields despite heavy leaf occlusion. Results show that the algorithm is able to detect 46% of the weeds in a field, despite the fact that large parts of the weeds overlap with wheat plants. The algorithm, however, encounters problems in detecting very small weeds, grasses, and weeds exposed to a severe degree of overlap. In addition, the algorithm for large plants has trouble generating optimal bounding boxes that cover the whole plant.

Acknowledgements

This work was founded by GUDP (Grønt Udviklings- og DemonstrationsProgram) under the Ministry of Environment and Food of Denmark.

References

Andújar D, Weis M and Gerhards R 2012. An ultrasonic system for weed detection in cereal crops. Sensors (Basel, Switzerland), 12 (12), 17343–17357.

Barker J, Sarathy S and Tao A 2016. "DetectNet: Deep Neural Network for Object Detection in DIGITS". *Nvidia*, (retrieved: 2016-11-30), https://devblogs.nvidia.com/parallelforall/detectnet-deep-neural-network-object-detection-digits/

Dyrmann M and Christiansen P 2014. "Automated Classification of Seedlings Using Computer Vision". Technical report, *Aarhus University*, Aarhus, Denmark.

Dyrmann M, Karstoft H and Midtiby HS 2016. "Plant Species Classification Using Deep Convolutional Neural Network." *Biosystems Engineering*.

Eurostat 2016, "Crop statistics (from 2000 onwards)", (retrieved: 2016-05-04), http://ec.europa.eu/eurostat/web/products-datasets/-/apro_acs_a

Giselsson TM 2014. "*Plant Object Classification in 2D Imagery*". PhD-thesis University of Southern Denmark, Denmark.

Herrmann I, Shapira U, Kinast S, Karnieli A and Bonfil DJ 2013. Ground-level hyperspectral imagery for detecting weeds in wheat fields. Precision Agriculture 14, 637–659.

Laursen MS, Jørgensen RN, Dyrmann M and Poulsen RN 2016. "RoboWeedSupport - Sub millimeter weed image acquisition in cereal crops with speeds up till 50 km/h", *European Conference of Precision Agriculture 2016*.

Laursen MS, Jørgensen RN, Midtiby HS, Jensen K, Christiansen MP, Giselsson TM and Jensen PK 2016. "Dicotyledon Weed Quantification Algorithm for Selective Herbicide Application in Maize Crops". Sensors 16 (11), 1848.

https://doi.org/10.1017/S2040470017000206 Published online by Cambridge University Press

Lottes P, Hoeferlin M, Sander S, Müter M, Schulze P, Stachniss LC and Stachniss C 2016. "An Effective Classification System for Separating Sugar Beets and Weeds for Precision Farming Applications". *Proc. of the IEEE International Conference on Robotics and Automation (ICRA) 2016*, 5157–5163.

Oquab M, Bottou L, Laptev I and Sivic J 2014. "Learning and Transferring Mid-level Image Representations Using Convolutional Neural Networks". *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 1717–1724.

Pérez AJ, López F, Benlloch JV and Christensen S 2000. Colour and shape analysis techniques for weed detection in cereal fields. Computers and Electronics in Agriculture 25 (3), 197–212.

Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC and Fei-Fei L. *ImageNet Large Scale Visual Recognition Challenge. IJCV,* 2015.

Rydahl P, Jensen N-P, Dyrmann M, Nielsen PH and Jørgensen RN 2016. "RoboWeedSupport - Presentation of a cloud based system bridging the gap between in-field weed inspections and decision support systems", *European Conference of Precision Agriculture 2016*.

Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D and Rabinovich A 2014. "Going Deeper with Convolutions". *arXiv Preprint* arXiv:1409.4842, 1–12.

https://doi.org/10.1017/S2040470017000206 Published online by Cambridge University Press