



Using a fully convolutional neural network for detecting locations of weeds in images from cereal fields

Mads Dyrmann¹, Søren Skovsen¹, Morten Stigaard Laursen¹, and Rasmus Nyholm Jørgensen¹

¹ Department of Engineering, Aarhus University, Denmark

**A paper from the Proceedings of the
14th International Conference on Precision Agriculture
June 24 – June 27, 2018
Montreal, Quebec, Canada**

Abstract. *Information about the presence of weeds in fields is important to decide on a weed control strategy. This is especially crucial in precision weed management, where the position of each plant is essential for conducting mechanical weed control or patch spraying.*

For detecting weeds, this study proposes a fully convolutional neural network, which detects weeds in images and classifies each one as either a monocot or dicot. The network has been trained on over 13 000 weed annotations in high-resolution RGB images from Danish wheat and rye fields. Due to occlusion in cereal fields, weeds can be partially hidden behind or touching the crops or other weeds, which the network handles.

The network can detect weeds with an average precision (AP) of 0.76. The weed detection network has been evaluated on an Nvidia Titan X, on which it is able to process a 5 MPx image in 0.02 s, making the method suitable for real-time field operation.

Keywords. *Weed detection, Deep-learning, Convolutional Neural Network, Weed control*

The authors are solely responsible for the content of this paper, which is not a refereed publication.. Citation of this work should state that it is from the Proceedings of the 14th International Conference on Precision Agriculture. EXAMPLE: Dyrmann et al., (2018). Using a fully convolutional neural network for detecting locations of weeds in images from cereal fields. In Proceedings of the 14th International Conference on Precision Agriculture (unpaginated, online). Monticello, IL: International Society of Precision Agriculture.



Fig. 1 Sample of partly hidden weeds in a wheat field. Image from training data (red=monocot, blue=dicot)

Introduction

In order to conduct site-specific weed management and utilize selective herbicides, it is necessary to know which weeds are present in the field. Furthermore, knowing weeds' locations can be utilized by variable dosage sprayers, thereby reducing the herbicide usage further, as herbicides are only allocated where needed. The composition of weed species can be inferred from images acquired in the field, which are analyzed automatically by a computer that then create optimal weed management strategies.

Cereal¹ fields account for 52% of the farmland in Northern Europe² (Eurostat, 2016). Despite this, cereal fields are often ignored in camera-based precision weed control as most research in automated precision weed control is targeted high-value crops with a row distance that allows for inspection between rows. Cereal fields are challenging as weeds are often only partially visible due to overlap from nearby crops and weeds, as illustrated in Figure 1. Furthermore, cereals are monocotyledons, they have many visual similarities with grasses, which further complicates detecting them.

After detecting weeds' locations, their species can be estimated and weed distribution maps created (Dyrmann, 2017) that can be converted to optimal spray plans (Rydahl et al., 2017). In the agricultural domain, various methods for weed detection has been proposed. Early attempts were based on color-segmentation, which is still seen used in row-crops, where the spatial location of plants can indicate whether plants are crops or weeds. However, in order to detect individual plant-instances in highly occluded fields such as cereal fields, it is necessary to distinguish plants even when they are only partly visible. Methods for detecting overlapping plants include one by Lottes et al., (2016), who demonstrate a method for detecting weeds in sugar beet fields by combining global and local texture features. The method is, however, designed for mechanical weed control, where it is sufficient to distinguish different species, but not instances of the same species. Likewise, Skovsen et al., (2017, 2018) have demonstrated a fully-convolutional network able to accurately distinguish clover from grass and weeds, but not individual instances.

Various convolutional neural network-based methods for object detection have been proposed in recent years. Best known is probably the RCNN method (Girshick et al., 2014), which contains a region proposer followed by a standard CNN for classification. RCNN and its derivatives fast-RCNN and faster-RCNN are accurate, but their architecture limits their usage due to long

¹ Barley, Durum wheat, Oats, Rye, Wheat and spelt, maslin, Triticale, and other cereals n.e.c. (buckwheat, millet, canary seed, etc.)

² Countries defined as Northern Europe by the United Nations Statistics Division: Denmark, Estonia, Finland, Iceland, Ireland, Latvia, Lithuania, Norway, Sweden and The United Kingdom

Table 1 Overview of images and weed instances in train and test-fields

Field	Training		Test		Crop (Latin)	Date
	# Images	# Plants	# Images	# Plants		
1	21	1341	0	0	Maize (<i>Zea mays</i>)	2016-06-14
2	479	2491	0	0	Winter rye (<i>Secale cereale</i>)	2017-04-04
3	842	7769	0	0	Winter wheat (<i>Triticum aestivum</i>)	2016-09-30
4	0	0	26	139	Winter wheat (<i>Triticum aestivum</i>)	2016-09-30
5	0	0	11	86	Winter wheat (<i>Triticum aestivum</i>)	2016-09-30
6	11	856	8	335	Winter wheat (<i>Triticum aestivum</i>)	2017-10-19
7	15	720	9	559	Winter wheat (<i>Triticum aestivum</i>)	2017-10-19
total	1368	13177	54	1119		

processing times on standard hardware. The YOLO-architecture (You Only Look Once)

eliminates the separated region proposer and classifier, by combining it in one fully-convolutional network, allowing it to run real-time on standard consumer hardware. The more recent SSD (Single Shot MultiBox Detector) (Liu et al., 2016) build upon the ideas from YOLO, but introduces a multiscale feature extractor, allowing objects to have more size-variation.

The present study is based on the weed detection methods described in our previous study (Dyrmann, Jørgensen, & Midtby, 2017), where a fully-convolutional network was used for detecting weed instances in cereal-fields. However, as that method was only trained to detect weeds, without distinguishing monocots from dicots, it had problems detecting both at the same time as it only included one detection-threshold; If the detection threshold was low enough to detect most dicots, cereals were also mistaken for monocots, and setting the threshold right for detecting monocots resulted in missed dicots.

Here, we base the weed detector on the Single Shot MultiBox Detector (SSD) by Liu et al., (2016), which utilizes the VGG16 network (Simonyan & Zisserman, 2014) as a multi-resolution feature extractor. The network is modified for detecting mono and dicotyledonous weeds in cereal fields despite occlusion.

Material and methods

In order to train the networks to be able to detect individual plants, it needs to be presented for a vast amount of training images to be able to handle real-world field variations.

Images for this study has been acquired in various fields and at different times of the growing season. Images from field no. 1 in Table 1, were acquired using a cellphone camera, and the rest were acquired with the automated camera system, mounted onto an ATV while driving at up to 50km/h (Laursen et al., 2017). This camera system consists of a Point Gray Grasshopper 3 that is synchronized with a xenon ring-flash. The camera exposes in 40µs per image, which ensures sharp images despite the speed. In the period from autumn 2016 till May 2018 the camera system carried by ATV has collected 118 000 images in winter cereal and spring cereal fields covering an approximate area of 770 hectares.

Images from maize, winter wheat and winter rye were collected from 7 fields. From five of those fields, 1368 images have been selected for training and validation, and 13 177 bounding boxes for plants have been annotated as either monocot or dicot.



Fig. 2 (a) Camera platform attached on the back of an ATV, (b) Sample image acquired at about 40km/h

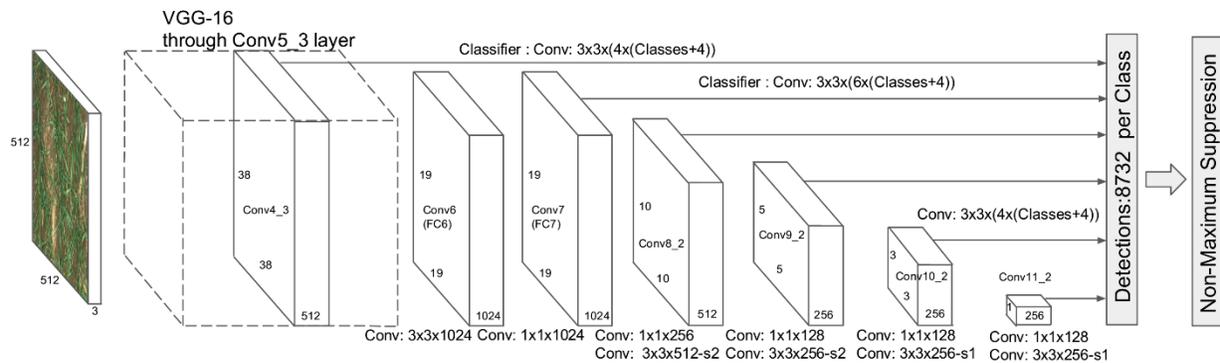


Fig. 3 Network architecture (after Liu et al., 2016)

Images from winter wheat and rye were collected from 7 fields. From five of those fields, 1368 images have been selected for training and validation, and 13 177 bounding boxes for plants have been annotated as either monocot or dicot.

Further 54 images have been selected for the test, in which 1119 plants have been annotated. The image-data is summed up in Table 1 and a sample is shown in Figure 2. The images have a ground sampling distance of 4-6px/mm.

Network architecture

The convolutional neural network used for training is based on the SSD512 architecture by Liu et al., (2016), which is a fully-convolutional network for object detection. As the network is fully-convolutional, it needs only a single forward-pass for determining plants' locations.

The network uses the first five layers of VGG16 as basic feature extraction layers, which ensures a rich pool of general features, that is combined in the network's latter layers. After these first five layers, sub sequential convolutional layers are added, which decrease in size progressively and thereby enables the network to detect plants at multiple scales.

In order to determine the locations of weeds, the network uses a set of default bounding boxes at each location in the final feature maps, for which it generates a score. The scores at these feature maps indicate the confidence in the presence of monocots or dicots at that location as well as how the default bounding boxes should be offset and stretched to fit the weeds' shapes in the images. The network architecture is sketched in Figure 3.

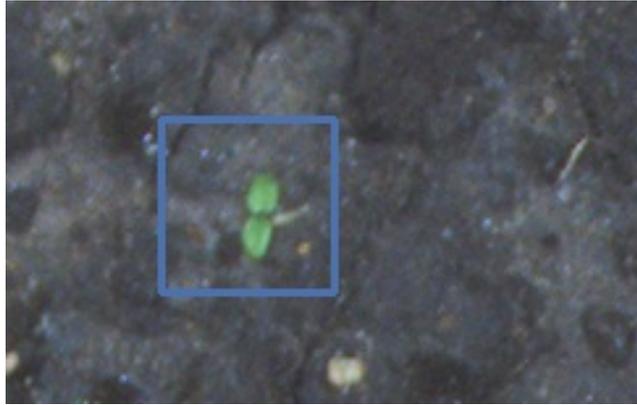


Fig. 4 Bounding boxes in training and test set is not fully aligned with weeds, which results in low intersections-over-union. In this sample, the intersection over union is 0.12

During training the images are cropped randomly between 6.25% and 100% of their size, and afterwards scaled to 512×512pixels. This scaling help making the network scale-invariant. Furthermore the hue was randomly altered to make the network invariant towards light-settings.

Training was conducted using the ADAM optimized with a cost that consists of a weighted sum between a localization loss and a detection loss. The localization loss is an L1 norm of the offsets of the predicted bounding boxes relative to the annotated bounding boxes. The detection loss is implemented as a softmax with cross-entropy loss, which indicates if an object is at the location of the bounding box.

Because of this latter part of the loss, all weeds in the training images must be annotated, as a missing plant in the training data will penalize the detector if it detects the unannotated weed. For further description and implementation details see (Liu et al., 2016) and (Balancap, 2017).

Evaluation

As to evaluate the weed detection ability, we assess the Intersection Over Union (IOU), which is the common area of a detected bounding box and an annotated bounding box in proportion to the area that the annotation and detection cover together.

A plant is considered detected if the IOU of the detected bounding box and the annotated bounding box is more than 0.10. Normally object detection algorithms are evaluated with a higher IOU, but the reason for this threshold is twofold: Firstly, the main objective is to count the number of weeds, for which reason precise bounding boxes is of less importance. Secondly, for small plants, the annotations are not fully aligned with the plants in the training and test data. This means that a detection that perfectly aligns with the plants, will not get an IOU of one. This is illustrated Figure 4, where a plant is shown with a manual annotation. In this case, the IOU is of a perfect detection and the manual annotation is only 0.12.

Results and discussion

A total of 1119 plants were annotated in the 54 test images. 682 of these 1119 annotated plants were detected automatically. This yielded a recall of 0.60 and a precision of 0.82. The average IOU for the accepted detections is 0.37. Except for the accepted detections, there are 152 detections, which does not match with any annotation. Overall, these false detections or missed detections can be categorized into five: cereal that is misclassified as weeds; small plants; plants exposed to heavy occlusion; plants where only part of the plant has been detected; and defoliation from nearby shelter-belts, that is confused with weeds.

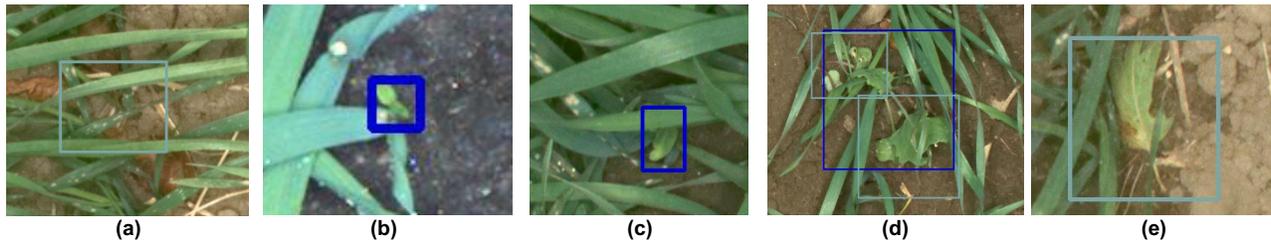


Fig. 5 Samples where the algorithm fails at detecting weeds. (a) cereal misclassified as weed, (b) undetected small weed, (c) undetected occluded weed, (d) large weed detected as multiple instances, (e) oak leaf misclassified as weed

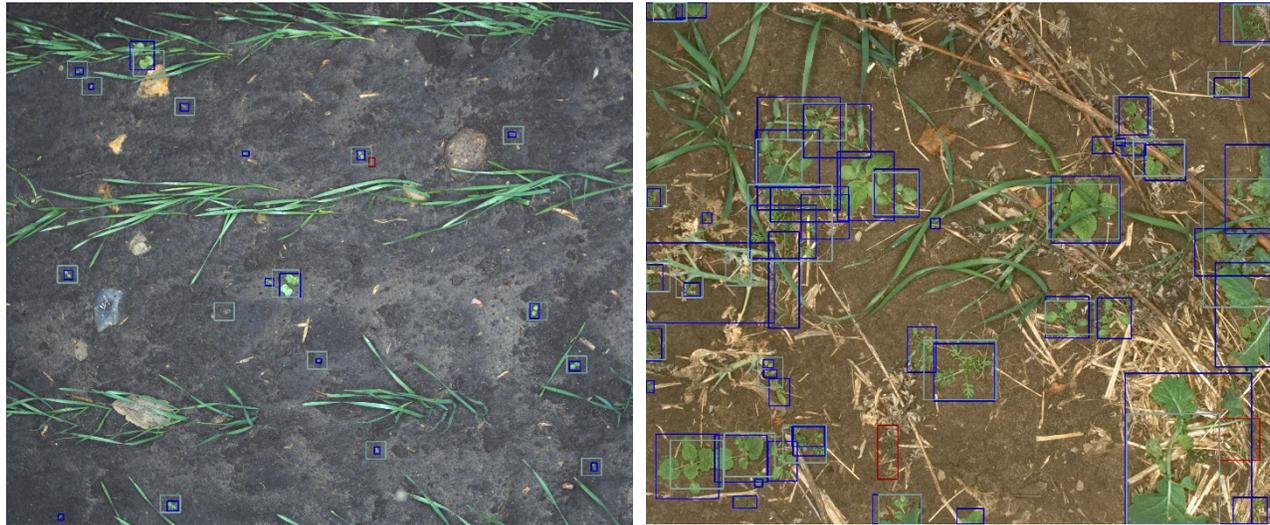


Fig. 6 Samples from the test dataset, where the algorithm's detections of dicots are shown in light blue, the ground truth dicots are shown in dark blue, and the ground truth monocots are shown in dark red (not detected)

The mean size of all the hand-annotated bounding boxes is 5137 pixels, ranging from 180 pixels for the smallest annotation up to 293 040 pixels for the largest annotation. An evaluation of the missed detections shows that on average the size of these bounding boxes is 2378 pixels, which indicates that small weeds are hardest to detect and that the network could benefit from more small weeds in the training data. Samples from the test images are shown in Figure 6, where the algorithm's suggested locations are shown along with the manual annotations.

The average precision of detected weeds is 0.76 i.e. without distinguishing monocots from dicots. However, when considering the mean average precision (mAP) across classes, it drops to 0.42. This is mainly due to monocots, which are less often detected and as they account for only 2.4% of all weeds, the influence from each monocot error is relatively larger. These results indicate that the method is suitable for detecting dicots in cereals, and to some extent monocots, which is, however, often mistaken for being cereals.

The images in the test set are relatively similar to those in the training set, and a decrease in detection accuracy should, therefore, be expected when evaluating on images that differ greatly from those in the training set. Nonetheless, it is believed that a more extensive training-set can make the method robust towards more variation in the images. Here it should be noted, that the network was trained several times without convergence before it succeeded. A larger training set will probably only make convergence harder and it is thus suggested to progressively add more variation to the training set, rather than training on all data from the start.

Conclusion

So far current methods for weed detection have had trouble with detecting weed instances in cereal fields due to heavy leaf occlusion.

We addressed the problem of detecting weed instances by using a modified version of the SSD detector, which is a fully convolutional neural network. This network has been trained to distinguish monocot weeds and dicot weeds from cereals. A previous attempt at detecting weeds using a fully convolutional neural network showed problems in handling both monocots and dicots simultaneously. This is sought solved by separating the monocot and dicot detector.

Results show that the algorithm achieves an average precision of 76% and that it detects 60.0% of the annotated weeds. However, large weeds cause trouble as they are often detected as multiple instances. Further training on images containing large weeds is, therefore, the natural next step.

References

- Balancap. (2017). SSD-Tensorflow. Retrieved from <https://github.com/balancap/SSD-Tensorflow>
- Dyrmann, M. (2017). *Automatic Detection and Classification of Weed Seedlings under Natural Light Conditions*. University of Southern Denmark.
- Dyrmann, M., Jørgensen, R. N., & Midtby, H. S. (2017). Detection of Weed Locations in Leaf Occluded Cereal Crops using a Fully Convolutional Neural Network. *11th European Conference on Precision Agriculture*, 842–847. <https://doi.org/10.13580/j.cnki.fstc.2016.03.008>
- Eurostat. (2016). Crop statistics (from 2000 onwards). Retrieved May 4, 2016, from http://ec.europa.eu/eurostat/web/products-datasets/-/apro_acs_a
- Girshick, R., Donahue, J., Darrell, T., Berkeley, U. C., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *Cvpr'14*, 2–9. <https://doi.org/10.1109/CVPR.2014.81>
- Laursen, M. S., Jørgensen, R. N., Dyrmann, M., & Poulsen, R. N. (2017). RoboWeedSupport - Sub millimeter weed image acquisition in cereal crops with speeds up till 50 km/h. *ICPA 2017 - 19th International Conference on Precision Agriculture*.
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., & Berg, A. C. (2016). SSD: Single Shot MultiBox Detector (pp. 21–37). https://doi.org/10.1007/978-3-319-46448-0_2
- Lottes, P., Hoferling, M., Sander, S., & Stachniss, C. (2016). Effective Vision-Based Classification for Separating Sugar Beets and Weeds for Precision Farming. *Journal of Field Robotics*.
- Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *ArXiv 2015*, 1–10. <https://doi.org/10.1016/j.nima.2015.05.028>
- Rydahl, P., Jensen, N.-P., Dyrmann, M., Nielsen, P. H., & Jørgensen, R. N. (2017). RoboWeedSupport - Presentation of a cloud based system bridging the gap between in-field weed inspections and decision support systems. *ECPA 2017 - 11th European Conference on Precision Agriculture*, 860–864. <https://doi.org/10.1017/S2040470017001054>
- Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv*. Retrieved from <http://arxiv.org/abs/1409.1556>
- Skovsen, S., Dyrmann, M., Mortensen, A. K., Steen, K. A., Green, O., Eriksen, J., Jørgensen, R. N., Karstoft, H. (2017). Estimation of the botanical composition of clover-grass leys from RGB images using data simulation and fully convolutional neural networks. *Sensors (Switzerland)*, 17(12). <https://doi.org/10.3390/s17122930>
- Skovsen, S., Dyrmann, M., Karstoft, H., Jørgensen, R. N., (2018). Predicting Dry Matter Composition of Grass Clover Leys Using Data Simulation and Camera-based Segmentation of Field Canopies into White Clover, Red Clover, Grass and Weeds, Proceedings, ICPA 2018