Active learning for efficient annotation in crop-weed semantic segmentation

Pieter M. Blok[1], Bart M. van Marrewijk[1], Dan Jeric Arcega Rustia[1,*], Charbel Dandjinou[2], Nicolas Franco Gonzalez[2], Paul Melki[2,3], Boubacar Diallo[2], and Jérôme Dias[2]

[1]Wageningen Plant Research, Wageningen University & Research, Wageningen, The Netherlands
[2]EXXACT Robotics, Épernay, France
[3]IMS, CNRS, University of Bordeaux, Bordeaux, France
[*]Corresponding author. Email: dan.rustia@wur.nl

The use of deep learning has received much attention in the agricultural domain [1]. Within deep learning, semantic segmentation plays an important role in accurately distinguishing class-specific pixels within agricultural images [2]. However, to achieve accurate segmentation performance, the neural network must be trained on a significant number of annotated images, a process that is time-consuming and expensive [3]. To address this challenge, active learning has emerged as a promising approach to optimize the annotation process by actively selecting the most informative images for annotation [4, 5]. Active learning has been applied on well-known public datasets such as Cityscapes, but for agricultural datasets, the amount of research is limited. This is unfortunate, because agricultural datasets have specific challenges, such as limited diversity and unbalanced classes. The aim of this paper is to test whether active learning can be applied to non-diverse and unbalanced agricultural datasets.

1. Active learning

Our proposed active learning framework is based on uncertainty sampling (BALD) [6] and hybrid uncertainty sampling (PowerBALD) [7]. Both methods select the images, the semantic segmentation network is most uncertain about. These images are assumed to contribute most to the performance when retraining the segmentation network. In this research, the uncertainty metrics were determined with Monte-Carlo dropout [8]. With Monte-Carlo dropout, unlabelled images are repeatedly passed through the neural network in inference mode with dropout activated, thus producing multiple segmentations for the same image. If the outputs for the same image show large variation, this may indicate that the model is uncertain about this image, making it a candidate for selection and annotation.

In our experiments the Fully Convolutional HarDNet (FCHarDNet) [9] was used as semantic segmentation model. To perform active learning, a dropout layer was added to the final convolutional layer of FCHarDNet. By adding this dropout layer, Monte-Carlo iterations could be performed. The number of Monte-Carlo iterations was set to 20. We used the default training parameters of FCHarDNet when conducting the active learning.

2. Experiments

Active learning was tested on two datasets: Cityscapes [10], and a proprietary dataset named Corn-Weed. The dataset statistics are shown in Table 1.

Dataset	Train	Validation	Test	# of classes	Majority class
Cityscapes	2975	500	1525	19	44.1% (road)
Corn-Weed	1190 (field A)			3	90.9% (background)
	331 (field B)	117 (field B)	-	3	

Table 1: Dataset statistical summary, including number of classes and pixel percentage of majority class.

The Cityscapes dataset [10] was used as a proof of concept to test our active learning. Cityscapes can be categorized as a diverse dataset, since the images were obtained from 18 different cities with non-overlapping frames. Furthermore, it consists of 19 classes, with the road class predominating with an average of 44% of pixels (Table 1). On Cityscapes, we compared three active learning acquisition functions: BALD, PowerBALD and Random selection. The number of active learning sampling iterations was set to 10. The initial dataset size was 59 images and the query size was 26 images. These numbers were obtained from a similar experiment in literature [11].

Our second dataset was a proprietary Corn-Weed dataset that consisted of 1638 images. The dataset consisted of two parts: part A & part B. Part A contained 1190 images from 17 fields spread over 3 countries. Part B was of a newly independent corn field and consisted of 448 images. In this experiment, the active learning performance was evaluated in an industrial application, meaning that a model was already trained on part A, but the performance of this model had to be improved for the new unseen corn field B. The question was which of the active learning acquisition functions selected the images most efficiently. Therefore, we made the comparison between BALD, PowerBALD and Random. Field B was subdivided into a validation set of 117 images, and the remaining 331 images were used as available images for active learning sampling (Table 1). The experiment was done with three repetitions to validate the stability of each acquisition function. The initial dataset size was 1190 images and the query size was 10 images. The total number of active learning sampling iterations was 10.

(a) Input image

(b) Ground truth annotation

Figure 1: Example image (a) and annotation (b) from Corn-Weed dataset. In image (b), red=corn, purple=weed and transparent=background.

3. Results

In Figure 2, the performance (mIoU) on the validation dataset of Cityscapes is shown as a function of the number of training images. From this figure, both BALD and PowerBALD have a higher mIoU than Random selection. This difference is already visible after the first sampling iteration, after which it is almost constant for the remaining iterations. In the last iteration, the mIoU value for BALD, PowerBALD and Random was 0.38, 0.38 and 0.36 respectively. The maximum mIoU of 0.36 when doing Random selection, was achieved after sampling 225 images when doing BALD and PowerBALD. This indicates that active learning is improving annotation efficiency by having a similar mIoU with 70 annotations less than Random sampling.

The result on the Corn-Weed dataset is shown in Figure 3. In the last iteration, the mIoU value for Random was 0.68, while the mIoU for BALD and PowerBALD was 0.71 and 0.72, respectively. This indicates that PowerBALD is performing best for active learning sampling. Additionally, PowerBALD proved to be more stable than the other methods, exhibiting lower variance across the active learning iterations, resulting in a significant different (p=0.01) between PowerBALD and Random sampling. Similar to Cityscapes, actie learning achieve the same mIoU than random requiring 70 less images.

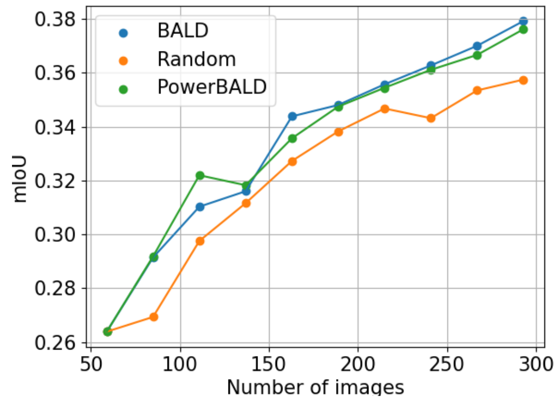

Figure 2: Validation performance (mIoU) as a function of the number of training images for BALD, PowerBALD, and Random on the Cityscapes dataset.

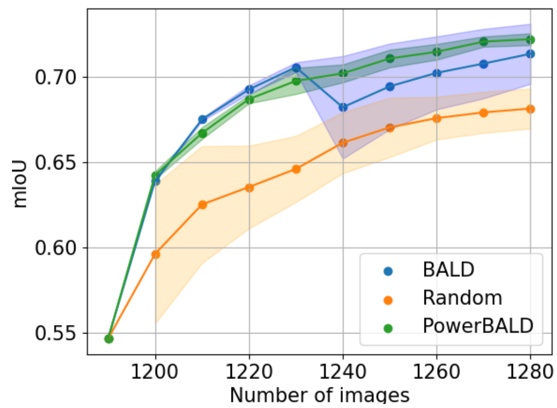

Figure 3: Validation performances for BALD, PowerBALD and Random acquisition on the Corn-Weed dataset. The transparent colored areas around the solid lines represent the 95% confidence interval for the three repetitions. The solid line is the average mIoU over the three repetions.

4. Conclusions

Active learning focuses on optimizing neural networks with fewer image annotations. For active learning on semantic segmentation networks, Cityscapes is a frequently used benchmark dataset. Because the Cityscapes dataset is quite diverse and consists of many classes, it does not accurately reflect active learning performance in agricultural settings. In agriculture, datasets tend to contain more redundant images and imbalanced classes. Therefore, in this research the added value of active learning was tested on a Corn-Weed dataset. Three acquisition functions were compared: BALD, PowerBALD and Random. Both BALD and PowerBALD outperformed Random sampling even when 90.9% of the pixels belonged to the background class. The results between PowerBALD and Random were significant showing that active learning can work in agriculture settings.

References

[1] Andreas Kamilaris and Francesc X. Prenafeta-Boldú. Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147:70–90, 2018. 1

[2] Zifei Luo, Wenzhu Yang, Yunfeng Yuan, Ruru Gou, and Xiaonan Li. Semantic segmentation of agricultural images: A survey. *Information Processing in Agriculture*, 2023. 1

[3] Jiajia Li, Dong Chen, Xinda Qi, Zhaojian Li, Yanbo Huang, Daniel Morris, and Xiaobo Tan. Label-efficient learning in agriculture: A comprehensive review, 2023. 1

[4] Yarin Gal, Riashat Islam, and Zoubin Ghahramani. Deep bayesian active learning with image data. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, page 1183–1192. JMLR.org, 2017. 1

[5] Pieter M Blok, Gert Kootstra, Hakim Elchaoui Elghor, Boubacar Diallo, Frits K van Evert, and Eldert J van Henten. Active learning with maskal reduces annotation effort for training mask r-cnn on a broccoli dataset with visually similar classes. *Computers and Electronics in Agriculture*, 197:106917, 2022. 1

[6] Neil Houlsby, Ferenc Huszár, Zoubin Ghahramani, and Máté Lengyel. Bayesian active learning for classification and preference learning. *arXiv preprint arXiv:1112.5745*, 2011. 1

[7] Andreas Kirsch, Sebastian Farquhar, Parmida Atighehchian, Andrew Jesson, Frederic Branchaud-Charron, and Yarin Gal. Stochastic batch acquisition for deep active learning, 2022. 1

[8] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, ICML'16, page 1050–1059. JMLR.org, 2016. 1

[9] Ping Chao, Chao-Yang Kao, Yushan Ruan, Chien-Hsiang Huang, and Youn-Long Lin. Hardnet: A low memory traffic network. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 3551–3560, 10 2019. 1

[10] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 1

[11] Sudhanshu Mittal, J Niemeijer, J Schäfer, and Thomas Brox. Best practices in active learning for semantic segmentation. *arXiv:2302.04075v2 [cs.CV]*, 2023. 1