AI-supported data annotation in the context of UAV-based weed detection in sugar beet fields using Deep Neural Networks

Jonas Boysen[1] and Anthony Stein[1]

Abstract: Recent Deep Learning-based Computer Vision methods proved quite successful in various tasks, also involving the classification, detection and segmentation of crop and weed plants with Convolutional Neural Networks (CNNs). Such solutions require a vast amount of labeled data. The annotation is a tedious and time-consuming task, which often constitutes a limiting factor in the Machine Learning process. In this work, an approach for an annotation pipeline for UAV-based images of sugar beet fields of BBCH-scale 12 to 17 is presented. For the creation of pixel-wise annotated data, we utilize a threshold-based method for the creation of a binary plant mask, a row detection based on Hough Transform and a lightweight CNN for the classification of small, cropped images. Our findings demonstrate that an increased image data annotation efficiency can be reached by using an AI approach already at the crucial Machine Learning-process step of training data collection.

Keywords: weed detection, data annotation, Convolutional Neural Networks, semantic segmentation, interactive AI

1 Introduction

The application of herbicides on agricultural fields and their impact on the environment are highly discussed in society. Site-specific spraying can reduce the applied amount of herbicides by between 14.0 % and 39.2 % in maize [Ca17]. Application maps enable site-specific spraying and can be computed from geo-referenced drone images [Fe18]. The computation of the application maps can be done by segmenting the images [Ca17].

Recent Deep Learning-based Computer Vision methods include image segmentation by utilizing Convolutional Neural Networks (CNNs), which are applied for semantic segmentation. Semantic segmentation is the pixel-wise assignment of classes and requires data-intensive training of the underlying CNN-models by means of pixel-wise annotated images [Al21]. Such a pixel-wise annotation of training data is a very time-consuming task [Be20]. In this work, an annotation pipeline for UAV-based images of sugar beet and weed plants is created with the goal to increase the annotation efficiency and quality, i.e., by reducing annotation errors in this crucial step.

[1] University of Hohenheim, Department of Artificial Intelligence in Agricultural Engineering, Garbenstr. 9, 70599 Stuttgart, Germany, jonas.boysen@uni-hohenheim.de, anthony.stein@uni-hohenheim.de

2 Related work

In agricultural fields, CNNs have successfully been used in several tasks including the semantic segmentation of crops and weed in canola fields [AB20] and rice, sugar beet and carrot images [Kh20]. Apart from the segmentation of different types of plants, also binary plant masks of images have been created by using threshold methods using the Excess Green Index (ExG) [Wo95] or combining different color spaces [RRG20; Ta20]. These binary plant masks have also been successfully created using CNNs which are applied for semantic segmentation [Fa19; Ta20].

The detection of rows in soy bean fields has been accomplished by Bah et al. [BHC18] through utilization of Hough Transform for the line detection in combination with a method to detect the main angle of the crops. CNNs have been used to classify small cropped images as crop or weed plants [MLS17; Fa19] as well as in specific plant species [Pe20]. In these works, the small images are cropped from the original images based on connected pixels on a binary plant mask.

3 Approach

In this work, RGB images with a size of 9504 x 6336 pixels displaying on average about 260 sugar beet plants in BBCH-scale from 12 to 17 and 310 weed plants are used. The images are sampled from a database of one of thirteen flights each. For the processing and annotation of the images, a software tool with a visual interface has been developed. Initially, a high level of manual handling is required, which however can be substantially reduced by automation in later iterations. One iteration here describes the processing of one of these full-sized drone images.

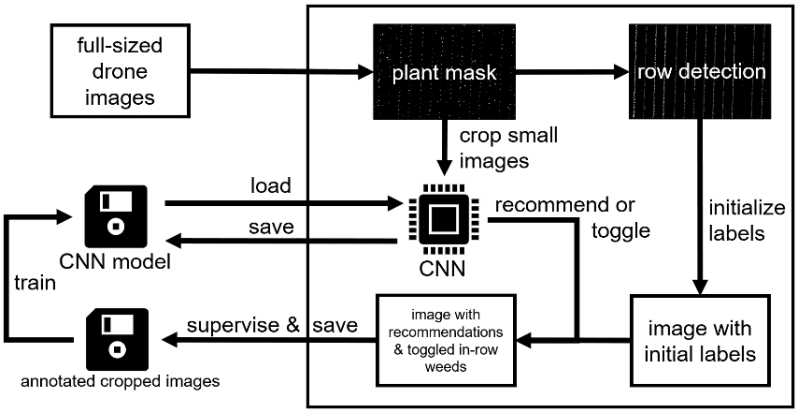

Fig. 1: The workflow of the annotation pipeline. This schematic illustrates the case when the annotator CNN is still trained after each iteration.

The whole workflow of the annotation pipeline is displayed in Figure 1. In the first step, the user of the software needs to create a binary plant mask, which is then utilized to find connected white pixels. These instances of connected white pixels are cropped with a bounding rectangle from the original images and are considered as the plants of the image. The plant mask is computed in two different ways in this work. The first approach to create the binary plant mask is threshold-based (TB). The method of Riehle et al. [RRG20] is used to propose a threshold that the user can manually adapt with visual feedback to create the binary plant mask. The second approach to create the binary plant mask is called the ground truth (GT) plant mask. It is similar to the first approach but includes a fine-tuning by the user who can improve the plant mask by drawing or erasing pixels.

In the next step, a row detection utilizing Hough Transform is performed to annotate all plants within rows as sugar beet plants and all plants between the rows as weeds. The crop rows are identified similar to the method proposed by Bah et al. [BHC18] by detecting lines on the plant mask. Subsequently, the main angle is calculated by putting the angles of all lines in a histogram and selecting the bin with the most members. The threshold of the Hough Transform-based line detection can be adapted until all rows are found.

After the initial annotation by the row detection, the user can supervise the annotation and toggle wrong labeled plants manually. When the annotation of the first image is completed, all plants are automatically cropped in a size of 64 x 64 pixels and saved with the respective label. In case the plants are of different size, smaller plants are centered in the bounding rectangle and larger plants are interpolated to fit the rectangle. The cropped images are used to train a classification CNN, which is called the annotator CNN. Subsequent iterations can use the annotator CNN to classify the plants found on the plant masks. The classifications of the annotator CNN are used to toggle detected weed plants that are positioned in the detected crop rows and give the user recommendations where other plants are predicted differently to the initialization of the row detection. This way the user can focus on the crop rows when manually supervising the annotations before saving them.

The combination of automatic plant annotation, annotation with human supervision and two different plant masks result in a total of four methods. These are tested for their performance in terms of the Intersection over Union[2] (IoU) and their speed recorded by the time of a single user who annotated all images. The tests utilize an annotator CNN which already has been trained on thirteen full-sized drone images resulting in about 10,000 cropped training images doubled by the use of data augmentation techniques [Al21]. The experiments include an evaluation regarding the errors due to the annotation process. Furthermore, the performance of CNN-models trained on the created data applied for semantic segmentation is evaluated in a ten-fold cross-validation manner. The CNN applied for semantic segmentation is a U-net [RFB15] with a pre-trained VGG-16 [SZ15] backbone using Transfer Learning and has an input size of 512 x 512 pixels. The annotator CNN is rather shallow as it comprises only four convolutional layers.

[2] Area of overlap divided by the area of union

4 Results and Discussion

For the evaluation of the four different methods, thirteen full-sized images are annotated with each method. All methods share the same annotator CNN-model pre-trained on a different set containing thirteen full-sized drone images. The masks created by the method using the GT plant mask and the annotation supervised by the user are considered the ground truth masks. In the following evaluation, each image is cut into 216 smaller images of size 512 x 512 pixels and the intersection over union (IoU) of the results with the ground truth masks is compared. The results of the comparison of the training data and the U-net are shown in Figure 2 in the respectively labelled rows. The rows display the mean IoU and the IoU of the single classes. The U-net predictions are performed by models which are trained on the training data created using the respective method. All statistical differences are calculated with a Kruskal-Wallis-Test since most of the data series are not normally distributed and are of homogeneous variances. Entries in the same row not sharing a letter are significantly different. The first row of the table displays the average time needed for the user to finish the annotation of one of the thirteen images.

	TB plant mask automatic labels	TB plant mask supervised labels	GT plant mask automatic labels	GT plant mask supervised labels
avg. time / img	1.7 min.	6.9 min.	59.8 min.	64.8 min.
avg. mean IoU ±1*SD*				
training data	72.6%^{a} ±5.36	77.1%^{a} ±6.45	88.8%^{b} ±7.00	(100%)
U-net pred.	75.9%^{a} ±0.64	77.0%^{a} ±0.59	79.6%^{b} ±1.03	79.3%^{b} ±0.70
avg. backg. IoU ±1*SD*				
training data	99.6%^{a} ±0.40	99.6%^{a} ±0.40	100%^{b} ±0.00	(100%)
U-net pred.	99.6%^{a} ±0.03	99.6%^{a} ±0.02	99.8%^{b} ±0.02	99.8%^{b} ±0.01
avg. s. beet IoU ±1*SD*				
training data	79.7%^{a} ±7.59	82.1%^{a} ±6.26	95.6%^{b} ±6.62	(100%)
U-net pred.	82.2%^{a} ±0.86	82.3%^{a} ±0.91	90.8%^{b} ±0.80	90.8%^{b} ±0.71
avg. weed IoU ±1*SD*				
training data	38.6%^{a} ±19.3	49.6%^{a} ±18.7	70.9%^{b} ±22.5	(100%)
U-net pred.	48.6%^{a} ±4.65	51.1%^{a} ±6.15	48.6%^{a} ±8.45	46.2%^{a} ±3.66

Fig. 2: Measured time needed for the user to annotate with either the TB or the GT plant mask with either automatic or supervised annotation of single plants and the performance of the annotation with n = 13 for the training data and n = 10 for the U-net predictions (±1 standard deviation, $\alpha = 5\%$).

The two methods using the TB plant masks are significantly different in the mean IoU as well as the IoU of every class to the method using the GT plant mask and automatic labels by the annotator CNN. No significant difference exists between the methods sharing the same plant mask but using different annotation methods. The more automated methods have on average a lower IoU than the methods using more manual techniques. It is observable that the weed class has the lowest average IoU values. The average values of the ten-fold cross-validation of the U-net predictions also show significant differences between the methods which use different plant masks except for the weed class. No significant differences are found in the weed class across all four methods. Methods that share the same plant mask also do not show significant differences.

A qualitative analysis including several samples shows that the TB plant mask interprets plant borders differently than the user does in the GT plant masks, as the TB mask applies less detailed and oversized borders. This explains the significant differences between the methods with different plant masks which are present in the created training data as well as in the predictions of the U-net except for the weed class. The usage of the annotator CNN to automate the annotations instead of giving recommendations and toggling only in the plant rows does not significantly change the IoU. Especially after the created training data is used for the training of the U-net, the average IoU of the methods sharing the same plant mask and using different annotation methods are less different.

Since the IoU of the weed class is not showing significant differences for all methods after using the training data for the training of U-net models, the choice of the method has no significant impact on the IoU of the important weed class. This result is confined by the fact that the IoU of the weed class is generally very low, indicating that the detection of the weed class is not learned well by the U-net. This is influenced by the class imbalance in the data set because the sugar beet class contains more pixels than the weed class.

The increased levels of automation reduce the average amount of time the user needs to create the training masks. The majority of time can be saved by using the TB plant mask instead of creating the GT plant mask. The automated annotations of the annotator CNN are especially useful when using the TB plant mask because a high amount of time relative to the total amount of time can be saved.

In summary, the iteratively increasing degree of automation and the interactive use of the annotator CNN in the proposed methods are expected to be of high potential value in practical use. They constitute an initial step towards a completely automatic annotation by incorporating a modified version of the U-net for the plant mask generation similar to Fawakherji et al. [Fa19] and combining this plant mask with an automatic annotator CNN.

5 Conclusion

The annotation process can be accelerated by using the methods of this work including the two different plant masks, the row detection and the different applications of the annotator CNN. The increased levels of automation reduced the time effort of the user and thus increased the annotation efficiency. Most of the errors in the created training data originate from the plant mask creation, while the annotator CNN showed no significant impact on the IoU. After the created training data has been used for the training of a U-net model, the differences in the IoU originating from the annotator CNN even reduced, but the differences between methods with different plant masks stayed significant.

Acknowledgment

Thanks to SAM-DIMENSION (https://sam-dimension.com) for providing the images used in this work and to Patrick Hansen for his contribution in the initial project.

References

[AB20] Asad, M. H.; Bais, A.: Weed detection in canola fields using maximum likelihood classification and deep convolutional neural network. Information Processing in Agriculture 7(4), 535-545, 2020.

[Al21] Alzubaidi, L. et al.: Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of Big Data 8(1), 1-74, 2021.

[Be20] Beck, M. A. et al.: An embedded system for the automated generation of labeled plant images to enable machine learning applications in agriculture. PLOS ONE 15(12), 1-23, 2020.

[BHC18] Bah, M.; Hafiane, A.; Canals, R.: Deep Learning with Unsupervised Data Labeling for Weed Detection in Line Crops in UAV Images. Remote Sensing 10(11), 1-22, 2018.

[Ca17] Castaldi, F. et al.: Assessing the potential of images from unmanned aerial vehicles (UAV) to support herbicide patch spraying in maize. Precision Agriculture 18(1), 76-94, 2017.

[Fa19] Fawakherji, M. et al.: Crop and Weeds Classification for Precision Agriculture Using Context-Independent Pixel-Wise Segmentation. In (IEEE): 2019 Third IEEE International Conference on Robotic Computing (IRC), 146-152, 2019.

[Fe18] Fernández-Quintanilla, C. et al.: Is the current state of the art of weed monitoring suitable for site-specific weed management in arable crops? Weed Research 58(4), 259-272, 2018.

[Kh20] Khan, A. et al.: CED-Net: Crops and Weeds Segmentation for Smart Farming Using a Small Cascaded Encoder-Decoder Architecture. Electronics 9(10), 1-16, 2020.

[MLS17] Milioto, A.; Lottes, P.; Stachniss, C.: Real-Time blob-wise Sugar Beets VS Weeds Classification for Monitoring Fields Using Convolutional Neural Networks. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. IV-2/W3, 41-48, 2017.

[Pe20] Peteinatos, G. G. et al.: Weed Identification in Maize, Sunflower, and Potatoes with the Aid of Convolutional Neural Networks. Remote Sensing 12(24), 1-22, 2020.

[RFB15] Ronneberger, O.; Fischer, P.; Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In (Springer): MICCAI, 234-241, 2015.

[RRG20] Riehle, D.; Reiser, D.; Griepentrog, H. W.: Robust index-based semantic plant/background segmentation for RGB- images. Computers and Electronics in Agriculture 169, 105201, 2020.

[SZ15] Simonyan, K.; Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556, 2015.

[Ta20] Tausen, M. et al.: Greenotyper: Image-Based Plant Phenotyping Using Distributed Computing and Deep Learning. Frontiers in Plant Science 11, 1-17 2020.

[Wo95] Woebbecke, D. M. et al.: Color indices for weed identification under various soil, residue, and lighting conditions. Transactions of the ASABE 38(1), 259-269, 1995.