Automatic UAV-based counting of seedlings in sugar-beet field and extension to maize and strawberry

October 4, 2021

Abel Barreto [a*], Philipp Lottes [b], Facundo Ramón Ispizua Yamati [a], Stephen Baumgarten [c], Nina Anastasia Wolf [c], Cyrill Stachniss [d], Anne-Katrin Mahlein[a] and Stefan Paulus [a]

Corresponding and main author: Abel Barreto; E-mail address: barreto@ifz-goettingen.de; Tel.:+49-551-505620

[a] Institute of Sugar Beet Research (IfZ), Holtenser Landstraße 77, 37079 Göttingen, Germany; barreto@ifz-goettingen.de (A.B.); ispizua@ifz-goettingen.de (F.I.); mahlein@ifz-goettingen.de (A.K.M.); paulus@ifz-goettingen.de

[b] Pheno-Inspect GmbH, Straßburger Straße 109, 46047 Oberhausen, Germany; philipp.lottes@phenoinspect.de (P.L.)

[c] ARGE NORD e.V., Helene-Künne-Allee 5, 38122 Braunschweig, Germany; stephen.baumgarten@arge-nord.de (S.B.); ninaanastasia.wolf@arge-nord.de (N.A.W)

[d] University of Bonn, Photogrammetry & Robotics Lab, Nussallee 15, Bonn

53115, Germany; cyrill.stachniss@igg.uni-bonn.de (C.S)

Abstract

Counting crop seedlings is a time-demanding activity involved in diverse agricultural practices like plant cultivating, experimental trials, plant breeding procedures, and weed control. Unmanned Aerial Vehicles (UAVs) carrying RGB cameras are novel tools for automatic field mapping, and the analysis of UAV images by deep learning methods can provide relevant agronomic information. UAV-based camera systems and a deep learning image analysis pipeline are implemented for a fully automated plant counting in sugar beet, maize, and strawberry fields in the present study. Five locations were monitored at different growth stages, and the crop number per plot was automatically predicted by using a fully convolutional network (FCN) pipeline. Our FCN-based approach is a single model for jointly determining both the exact stem location of crop and weed plants and a pixel-wise plant classification considering crop, weed, and soil. To determinate the approach performance, predicted crop counting was compared to visually assessed ground truth data. Results show that UAV-based counting of sugar-beet plants delivers forecast errors lower than 4.6%, and the main factors for performance are related to the intra-row distance and the growth stage. The pipeline's extension to other crops is possible; the errors of the predictions are lower than 4% under practical field conditions for maize and strawberry fields. This work highlight the feasibility of automatic crop counting, which can reduce manual effort to the farmers.

Keywords: deep learning, FCN, UAV, sugar beet, plant segmentation, time-series, intra-row distance, growth stage

1 Introduction

The counting of sugar-beet seedlings is a time-demanding activity necessary in diverse aspects of beet production such as plant cultivating, experimental trials, plant breeding, plant phenotyping, and weed control. At the beginning of the cultivation season and few days after sowing (das), plant population and location becomes a relevant parameter to describe plant distribution's homogeneity on the field.

Previous studies highlighted the relevance of plant density in final beet yield and quality, as well as for the white sugar yield (WSY; Märländer, 1990). Sub-optimal plant distribution can be caused not only by biotic factors like insects or fungi but also by abiotic factors like drought, crust formation, hail, wind, or frost damage (Smit, 1993). In Germany, farmers determine plant population by an intensive manual counting of sugar-beet seedlings at a growing stage (BBCH) 10-12 in a previously defined and representative patches of 10 m^2, the average value of all counts are extrapolated to the entire field. In Europe, the counted number of emerged plants defines the decision of resowing, where fields with a population lower than 45 thousand plants per hectare are decided to be replanted. Moreover, plant populations between 82 and 110 thousand plants per hectare are considered as optimal (Märländer, 1990), where seeds are frequently sown with a row distance of 45 and 50 cm and an intra-row distance from 18 to 25 cm. In experimental fields, a narrow sowing method is used where seeds are sown with a intra-row distance from 4 to 16 cm (Durrant et al., 1985; Pospišil et al., 2000; Söğüt and Arioglu, 2004) to achieve two or three times the population number employed in the practice, later young plants are commonly thinned out to manually achieve the optimal population.

Automatic counting of sugar-beet plants presents a potential for experi-

mental fields. In the European Union (EU), the demand for improvements in the sugar-beet cultivation to secure the crop yield and fulfill the principles of integrated pest management (IPM) leads to constant evaluations of new sugar-beet varieties, fungicides, insecticides and herbicides. Unlike counting made by sugar-beet growers, counting in field experiments varies according to experimental aims and requires a precise number of emerged plants per plot. Manual counting of field trials could amount from 120 to 250 person-hours per counted hectare, and counted area can represent 20 to 50% of the entire experimental field. Plant breeding trials also require the detailed monitoring of plant emergence, germination capacity of seeds under the field conditions is a time-demanding key parameter to quantify seed quality. Emergence rate, together with vigorous seedlings, determine if a new sugar-beet variety is accepted or rejected by the market (Milošević et al., 2010). Finally, competition for light and nutrients between uncontrolled weed and sugar beet can cause root yield losses from up to 95% (Petersen, 2004); therefore, modern post-emergence weed control activities like weeding through tractor-mounted hoes or herbicide application in weed-infested areas demands techniques for precise weed detection, location and distribution on the field to ensure crop yield and avoid the environmental impact of intensive herbicide application (Cioni and Maines, 2010; Kunz et al., 2015).

Unmanned Aerial Vehicles carrying RGB cameras appear as a novel automatic approach to sequential images for field mapping. Advantages for plant phenotyping are related to ease of operation, high spatial resolution, and acquisition of data on demand (Deng et al., 2018). RGB and RGB-NIR images were used in the past to detect sugar beet and weeds. The image processing pipeline includes the use of an end-to-end trainable fully convolutional network acquired from field robot platforms (Lottes et al., 2018b; Wu et al., 2020; McCool et al., 2017). This algorithm can

deliver information to track crops and weeds, including the number of plants, their location, and distribution on the field. The FCN is also able to detect stem position of sugar beets and weeds from RGB images with a ground sampling distance (GSD) of 2 mm (Lottes et al., 2018a). To emulate the complexity of field conditions, an improvement of this algorithm approach was also developed for field robot platforms and RGB-NIR images with 1 mm GSD to detect sugar-beet and weed stems under varying weed pressure, various weed types, and for different locations and growth stages of 2-, 4-leaf and 6-to-8-leaf (Lottes et al., 2019). Nevertheless, and although mentioned advantages of the image processing approach, there is still a lack of knowledge in following aspects: (1) the possibility to extend the approach to UAV-RGB systems and its robustness in different locations and in time-series, especially because UAV-systems in comparison with field robots deliver, due to distance camera-object, higher spatial resolution and output images quality depends of the sunlight conditions; (2) behavior of forecast performance for crop counting between different growth stages; and (3) the effect of different intra-row distances in the forecast performance of crop counting, considering the most demanding counting activities for experimental fields but also for practice.

Besides, the performance of UAV-recorded RGB images and a FCN pipeline to determine the plant number in other crops is also unknown but valuable for practitioners to predict the yield. RGB images and deep-learning approaches were already reported to be useful for counting diverse plant species like rice plants and acacia trees (Lu et al., 2021; Tong et al., 2021). The use of a previously trained model to other crop species has to consider differences in shape and color which could decrease the robustness of the methodology. Despite this disadvantage, two points might be beneficial in the extension of the method to other crops: (1) the reliability of the plant identification step (plant-soil segmentation), and (2) the

ability of previously trained FCN based approach to differentiate diverse weeds of various botanical families (crop-weed discrimination). Whereas the sugar-beet experimental fields include natural weed development, it is proposed to evaluate a practical approach close to the productive application on strawberry and maize fields. In this study, the described FCN pipeline, which is modeled for sugar beet will be adapted and applied to strawberries, and maize for crop counting.

The general objectives of this study are: (1) evaluate the performance of forecasting the number of sugar-beet plants on the field by using a FCN-based approach and RGB images of mapped fields with two different intra-row distances, (2) determinate the best performance of the approach comparing different growing stages and in time-series and (3) evaluate the possibility of extending the methodology established for sugar-beets to strawberry and maize.

2 Materials and methods

2.1 Experimental fields

Location 1 - sugar-beet variety trial This field presented a simplified sugar-beet variety trial design, and was located in Göttingen, Lower Saxony, Germany. The design showed 24 plots arranged in six rows, each plot with a size of 2.7 m $\times$ 8.0 m. The trial field presented two sugar-beet varieties with different leaf orientation, a planophile and an erectophile. Seeds were sowed on the 9th of April in 2019; the initial intra-row distance was 6.8 cm, and the distance between row was 45 cm. Two days before the UAV-measurement, plant population was reduced to $1/3$ in order to increase intra-row to 21 cm. The soil type was Chromic Luvisol.

Location 2 - proof of concept A second trial was located in Börßum, Lower Saxony, Germany (Figure 1a). The design was structured into 84 plots with a size of 1.35 m $\times$ 7 m. The trial field presented 21 sugar-beet varieties. Seeds were sown on the 10th of April in 2019, with an original intra-row distance of 6.8 cm and 45 cm between rows. Previous to UAV-measurement, plant population was reduced as location 1 to 21 cm. The soil type was Chromic Luvisol.

Monitoring and preliminary evaluation of data from locations 1 and 2 were performed as a background for a "proof of concept" and support the design for a time-series monitoring to determinate crop population under "harsh weed conditions."

Location 3 - time-series monitoring in harsh weed conditions for sugar beet The field trial for weed control was located in Höckelheim, Lower Saxony, Germany. The design was originally structured into 40 plots with a size of 1.35 m $\times$ 7 m. The sowing date was the 15th of August 2019, and the initial intra-row distance was 6.8 cm and 45 cm between rows. Previous to UAV-measurement, plant population was reduced to $1/2$. No herbicides were applied to this field. The soil type was Chromic Luvisol.

Location 4 and 5 - maize and strawberry These fields were located in Klein-Altendorf, North Rhine-Westphalia, Germany. The maize dataset (location 4) had originally no plot structure (Figure 1b). Therefore, 20 images were extracted from a single row, each representing a plot with a size of 0.56 m $\times$ 28.9 m. The dataset contains almost no weeds and reflects maize plants in its post-emergence growth stage. However, some of the maize plants were already overlapped. The strawberry dataset (location 5) presents also no plot structure. Thus, we define 40 squared

regions of 9 m × 9 m. The dataset also contains small weeds that are located between and within the rows of strawberry plants. The soil type for both trials was Haplic Luvisol.

2.2 UAV monitoring systems

Three UAV-monitoring systems were used for the present study. Technical specifications are described in table 1.

Table 1: **Technical specifications** of our UAV-based monitoring systems.

setup	*UAV*	*camera*	*manufacturer*	*sensor size*	*resolution*	*focal length*
1	Inspire 2	Zenmuse X5S	DJI	4/3-inch	5280×3956 px	45 mm
2	Phantom 4 RTK	gimbal-attached	DJI	1-inch	5472×3648 px	8.8 mm
3	Phantom 4 PRO	gimbal-attached	DJI	1-inch	5472×3648 px	8.8 mm

2.3 UAV-Monitoring campaign and flight planning

The monitoring campaign started when plants achieved the growing stages between BBCH 12 and 16 for sugar beets, and BBCH 13-17 and BBCH 13-16 in the case of the maize and strawberry fields, respectively. Flights were performed in all locations in on one occasion with the exception of location 3 (time-series monitoring), which was monitored on three occasions (Table 2). For this location, ground control points (GCPs) were installed in the field corner points. In all locations, flights were performed within three hours of local solar noon. Flight mission was established using the software UgCS (SPH Engineering, Riga, Latvia), and the flight time was between 12 and 22 minutes. RGB images were captured in the photo mode (single shot) by distance with a shutter speed between 1/1400 s and 1/1000 s, and a sidelap/frontlap of 80%. The resolution

in GSD was fitted to field size between 1.5 and 3 mm (Table 2). Image stitching to an orthomosaic was performed using AgiSoft MetaShape (Agisoft LLC, St. Petersburg, Russia).

Sky conditions of each UAV-measurement was categorized in four classes: (1) clear, (2) low cloudy, (3) middle cloudy, and (4) cloudy; according to cloud covers from 0 to 33, 34 to 65, 66 to 95, and 96 to 100% of effective cloud amounts (Schreiner et al., 1993).

Table 2: **Trial fields**, location, crop, growing stage and flight specifications

No.	*location*	*GPS-coordinates*	*crop*	*BBCH*	*flight date*	*sky*	*setup*	*height*	*GSD*
1	(1) Göttingen	$51°33N\ 9°53E$	sugar beet	14-16	22.05.2019	cloudy	1	40 m	3.0 mm
2	(2) Börßum	$52°40N\ 10°35E$	sugar beet	16-18	28.05.2019	middle cloudy	1	40 m	3.0 mm
3				12	03.09.2019	cloudy	2	6 m	1.5 mm
4	(3) Höckelheim	$51°42N\ 9°57E$	sugar beet	12-14	11.09.2019	clear	2	6 m	1.5 mm
5				14-16	20.09.2019	low cloudy	2	6 m	1.5 mm
6	(4) Klein-Altendorf	$50°36N\ 6°59E$	maize	13-17	25.05.2018	low cloudy	3	10 m	3.0 mm
7	(5) Klein-Altendorf	$50°36N\ 6°59E$	strawberry	13-16	28.06.2018	low cloudy	1	40 m	3.0 mm

2.4 Collecting ground truth

Ground truth object (GT) was defined as the number of crops counted per plot. Crops were labeled and counted manually in images for each plot from the segmented orthomosaic image. In total, 24, 84, 120, 20, and 40 plots were counted for locations 1-5, respectively. In the case of location 3, the counts were done separately for each flight date (Table 2).

2.5 Estimation of crop density and distance

Crop density (CD) in crops per m^2 was calculated per plot by using the ground truth counts. Because the size of the plots in each location was the same but different between locations, the average GT (counted crops)

and the average CD was calculated per location. Intra-row distance (D) was calculated by using GT and the number of rows and row length of each plot.

2.6 Fully convolutional network-based plant classification

In this publication, the principal axis of the young shoot of sugar-beet seedling will be consider as stem. Keeping this clarification in mind, this section describes our FCN-based approach for joint plant classification and stem detection, which is based on a previous publication which was implemented in Python (Keras) (Lottes et al., 2018a). This system provides two outputs simultaneously. First, a plant mask represents a pixel-wise classification considering the classes crop, weed, and background (mostly soil). Second, the stems' positions for the detected crop and weed objects are represented by a stem mask. The approach's key architectural design feature is that the network shares the encoded features for classifying the stem regions and the pixel-wise classification using one encoder network and two task-specific decoder networks.

The processing pipeline executes the following key steps and is illustrated in Figure 2. First, each image was preprocessed. Next, the preprocessed images were fed into the one-encoder-two-decoder structured fully convolutional network. Outputs first include a per-pixel probability distribution $P(\omega^{plant} \mid \mathrm{Z})$ for describing the plant classification over the desired class labels $\omega^{plant} \in \{\mathrm{crop}, \mathrm{weed}, \mathrm{soil}\}$ for each observed pixel (Z); second, a per-pixel probability distribution $P(\omega^{stem} \mid \mathrm{Z})$ with $\omega^{stem} \in \{\mathrm{crop}, \mathrm{weed}, \mathrm{soil}\}$ representing regions within the image, which correspond to crop and weed stems. The label for each pixel is determined as the one with the highest probability by:

$$\omega^* = \underset{\omega}{\operatorname{argmax}}\, P(\omega^{plant} \mid \mathsf{Z}). \tag{1}$$

Finally, pixel-accurate stem positions were extracted, i.e., the stem mask, through a post-processing step, which will be described in Section 2.7.

2.6.1 Image preprocessing

The preprocessing step was applied to image patches obtained from orthomosaics. These images presented a width and height dimension of 512 pixels, and 1 mm of GSD after downsampling. To deliver high performance of the classifiers the different input data was preprocessed. In preprocessing steps, transformations were applied to the data to reduce its complexity and standardize it to some degree, increasing the chance that the machine-learning algorithm can provide better performance than without preprocessing it. Technically speaking, preprocessing can improve a classification system's generalization capabilities by aligning the training and test data distribution. The preprocessing was performed independently for each image and separately on all channels, i.e., red, green, blue. First, noise from each channel was removed by performing a blurring operation using a $[5 \times 5]$ Gaussian kernel given by the standard normal distribution, i.e., $\mu = 0$ and $\sigma^2 = 1$. Second, each image channel was standardized by its mean and standard deviation, respectively. Third, contrast stretch of the intensities to the range $[-0.5, 0.5]$ was performed, which implies a zero-centering of the data.

2.6.2 One-encoder two-decoder network architecture

Figure 2 depicts the proposed architecture of our joint plant and stem detection approach. This approach's main processing steps are the pre-

processing (red), the encoder, the plant decoder, the stem decoder, and the stem extraction (brown).

Two separate feature volumes were generated from the encoded and compressed visual code, one specialized for pixel-wise plant classification, and another for stem detection. Thus, two task-specific decoders were obtained, which perform an upsampling using a stridden transpose convolution (Dumoulin and Visin, 2016) with $[2 \times 2]$ kernel and a stride of 2. Both decoders also use dense blocks as their main building blocks and follow the same architectural design to produce the plant features and stem features. Moreover, both task-specific decoders use feature maps produced by the encoder through skip connections. The corresponding feature maps was concatenated by sharing the same spatial resolution from the encoder before using dense blocks for feature computation. Skip connections from the encoder to the decoders facilitate the recovery of spatial information (Badrinarayanan et al., 2015). Finally, the feature maps produced by the stem decoder and the plant decoder was transformed into the pixel-wise probability distribution over their respective class labels by a $[1 \times 1]$ convolution followed by a softmax layer to obtain $P(\omega^{plant} \mid \mathrm{Z})$ and $P(\omega^{stem} \mid \mathrm{Z})$. Note that It was tried to predict the area of the stem instead of regressing the stem location. This is key to use the same architecture for learning plant classification and stem locations.

2.6.3 FCN training

For learning, an NVIDIA 2080 TI card with 11 GB of VRAM was part of the equipment for our experiment. A dataset of 500 RGB images were used in a ratio of 75%, 5% and 20%, from these 375 images were considered for training and 25 images for validation as mentioned in Lottes et al. (2018b) and Lottes et al. (2018a). Additional 100 images were used as testing dataset. For better results, all datasets were diversified according

image conditions by considering weed pressure, illumination, and growing stage. A multi-task loss $\mathcal{L}$ was used combining the loss for the plant segmentation $\mathcal{L}_{\text{plant}}$ and the loss for the stem region segmentation $\mathcal{L}_{\text{stem}}$ as follows:

$$\mathcal{L} = (1 - \alpha)\mathcal{L}_{\text{stem}} + \alpha\mathcal{L}_{\text{plant}}, \tag{2}$$

where $\alpha = 0.5$. The loss $\mathcal{L}_{\text{plant}}$ is the weighted cross-entropy, where errors regarding the crop plants and weeds were penalized by a factor of 10. The loss $\mathcal{L}_{\text{stem}}$ is based on an approximation of the intersection over union (IoU) metric, as it is more stable with imbalanced class labels (Rahman and Wang, 2016), which is the case in our problem with under-represented stems as compared to the amount of soil. The multi-task loss also enables the sharing of information for learning the encoder, which can use the loss information from both decoders in the backward pass of the backpropagation. The stem locations were encoded as blobs with a diameter of $10\,\text{mm}$ in object space for training.

2.7 Stem Extraction

Given the probability distribution $P(\omega^{stem} \mid \mathsf{Z})$ encoding regions within the image, which correspond to crop stems and weed stems, a well-defined stem detection by a specific pixel location for the crop and weeds was desired. To this end, it was firstly calculated the stem mask according to Equation 1 by selecting the class with the highest label probability for each pixel. Next, the connected components $\mathcal{O}_j^\omega$ was determined for the crop and weed class and computed the weighted mean $\bar{\mathbf{x}}_j^\omega$ of the pixel locations by:

$$\bar{\mathbf{x}}_j^\omega = \frac{\sum_{\mathbf{x} \in \mathcal{O}_j^\omega} P(\omega = \omega \mid \mathbf{x}) \cdot \mathbf{x}}{\sum_{\mathbf{x} \in \mathcal{O}_j^\omega} P(\omega = \omega \mid \mathbf{x})}, \tag{3}$$

with ω = crop,weed. The weighted means $\bar{\mathbf{x}}_j^{\omega}$ for class c are then the stem detections that are reported by our approach.

2.7.1 FCNs applicability for crop counting using adapted transfer learning

Trained data based on a from previous work for sugar beets was used and can therefore realize a stable primary classifier. The classifiers' adaptation to strawberries and maize was made with additional training data from earlier projects in a supervised transfer learning manner, i.e., through a re-training of the model with further training data that includes strawberries and maize samples. The dataset used for training individual classifier for strawberry and maize, consisted of 130 RGB images per crop, considering from this total 120 images for training and 10 images for testing with a resolution of 512 $\times$ 512 pixels (Sec. 2.6.1). No validation set was used in this transfer learning stage because it was tried to minimize modifications in hyperparameter settings. For processing the test dataset, patch images presented a resolution of 2048 $\times$ 2048 pixels, and the processing-time per image was around 0.4 seconds with the hardware mentioned on Sec. 2.6.3.

2.8 Evaluation Metrics

After checking the absence of outliers, normality of variables, linearity, and homoscedasticity of each location, the Pearson Product-Moment Correlation Coefficient (PPMC) was calculated to determine the degree of linear correlation and whether significant differences exist between prediction (P) and GT in terms of counted crops.

Forecast error of pipeline approach was determined by calculating the mean bias error (MBE), mean absolute error (MAE), and mean absolute percentage error (MAPE) (Shcherbakov et al., 2013; Kato, 2016). For a

better agricultural interpretation, MAE was calculated per unit of evaluated area (MAEA) and the area for unitary predicted error is given as A_e. The systematic error to under- or over-forecast of the pipeline approach was evaluated by using the MBE parameter, this was defined by equation 4:

$$MBE = \frac{1}{N}\sum_{i=1}^{N}(P_i - GT_i) \tag{4}$$

where:

- N = total number of forecast events
- i = forecast event for plant counting in a plot
- GT_i = ground truth count
- P_i = prediction count

The magnitude of the forecast error was determined by the MAE of the predicted event given by:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|P_i - GT_i| \tag{5}$$

Based on the ground truth plant count, a percentage error of all forecast events was quantified using the MAPE of predicted numbers of crop plants by the following eq. 6:

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}100 \times |\frac{GT_i - P_i}{GT_i}| \tag{6}$$

By considering MAE as a scaled magnitude of forecast error of events and as a forecast event is applied in a specific area of the field, the MAEA expresses the number of crop units, which are under- or over-counted per unit area (Eq. 7):

$$MAEA = \frac{MAE}{a} \tag{7}$$

where:

- $MAE =$ mean absolute error expressed in counted crop units
- $a =$ area of forecast event

The area for unitary forecast error (Ae) is derived from MAEA. It is interpreted as the area to under- or over-count a crop by the employed pipeline approach (Eq. 8):

$$A_e = \frac{1}{MAEA} \tag{8}$$

3 Results

For all locations the average of counted and predicted crops presented a strong and significant correlation (Table 3).

3.1 Proof of concept assessments

The average of counted and predicted plant numbers were 232.3 and 228.1, and 95.9 and 95.8 plants for location 1 and 2, respectively (Table 3). The number of counted plants allows determining the CD values for locations 1 and 2, 10.75 and 10.15 plants/m^2. The average intra-row distance calculated was 21 and 22 cm, respectively, for locations 1 and 2.

The use of FCN pipeline delivers a negative MBE value in both locations or tends to under-count sugar-beet plants (Table 3). In location 1, MAE's magnitude is four sugar-beet plants in a plot area of 21.6 m^2 and with plants at BBCH 14-16. According to the definition of MAEA and Ae, in location 1 and at BBCH 14-16; for every 4.8 m^2, one sugar-beet plant is wrongly considered in the count. In location 2, the pipeline presents the lowest MAE and the highest correlation (PPMC=0.95), the magnitude for the count error is less than a plant for every plot area of 9.5 m^2 at BBCH 16-18. The Ae value shows that one sugar-beet plant is

wrongly counted every 26.0 m^2 under field arrangement conditions of location 2 and at BBCH 16-18. The lowest MAPE value is shown between both locations with the growing stage of BBCH 16-18 in sugar beet.

3.2 Low intra-row distance in time series

The time-series assessment of location 3 delivers the counted number of crop plants per plot between 181.6 and 185.4, with the highest plant number at 17 das. The predicted average number of crops presents a range from 180.8 to 181.4 plants per plot, with the highest amount of counted plants at 34 das (Table 3). The average value of crop density in the trial decreased from 19.6 to 19.2 plants/m^2 through the three measurements.

Over the monitoring period, predictions present negative MBE values or a tendency to under-count crop plants similarly to locations 1 and 2 (Table 3). The magnitude of the error count range between 7 and 8 sugar beets per plot. The Ae value range between 1.3 and 1.5 m^2. Either the smallest error or the highest degree of correlation was found at BBCH 14-16 with a MAPE value of 3.96% and a PPMC of 0.83.

Sugar-beet plants presented an average soil cover of 1.14 $\pm$ 0.31%, 1.95 $\pm$ 0.67%, and 2.90 $\pm$ 0.88% at 17, 25 and 34 das (Figure 5). Furthermore, weeds covered the soil with average values of 1.05%, 1.19%, and 1.80% at the previously mentioned dates, and the respective standard deviations were 0.03%, 0.26%, and 0.46%.

3.3 Generalized application of trained FCN to maize and strawberry

The average of counted and predicted crop numbers for the maize and strawberry field was 153.6 and 157.4, 29.1 and 30.2 plants, respectively (Table 3). The CD value for the maize field was 9.49 plants/m^2, and in

Table 3: **experimental results for crop counting**, calculation of plant distribution parameters,and error metrics based on predicted crop numbers for sugar beet, maize and strawberry fields

trial configuration							*prediction & calculations*			*metrics*					
location	crop	PN	das	BBCH	GSD	GT	P	PD^a	D^a	MBE	MAE	MAEA	Ae	MAPE	PPMC
1	sugar beet	24	43	14-16	3.0	232.3	228.1	10.75	0.21	-4.21	4.46	0.21	4.84	1.92	0.86 *
2	sugar beet	84	48	16-18	3.0	95.9	95.8	10.15	0.22	-0.08	0.42	0.04	26.04	0.43	0.95 *
	sugar beet		17	12	1.5	185.4	180.8	19.62	0.11	-4.65	8.40	0.77	1.29	4.57	0.82 *
3	sugar beet	40	25	14-16	1.5	182.0	181.3	19.26	0.12	-0.70	7.10	0.65	1.53	3.96	0.83 *
	sugar beet		34	16-18	1.5	181.6	181.4	19.22	0.12	-0.18	8.03	0.74	1.35	4.44	0.78 *
4	maize	20	-	14- 16	3.0	153.6	157.4	9.49	0.19	3.80	3.90	0.24	4.15	2.55	0.92 *
5	strawberry	40	-	16	3.0	29.1	30.2	0.36	0.31	1.05	1.15	0.13	7.83	3.96	0.67 *

where: PN= plot number, das= days after sowing, BBCH= growing stage, GSD= ground sample distance in mm, GT= average value of ground truth expressed in number of plants, P= average value of predictions expressed in number of plants, PD= plant density in plants per m^2, D= distance of crops in the row in m considering a row distance of 45 cm for sugar beets, and 56 and 100 cm for maize and strawberry respectively, MBE= mean bias error, MAE= mean absolute error in $plant/plot$, MAEA= mean absolute error of predicted area in plants per m^2, Ae= area for unitary forecast error in m^2 per plant, MAPE= mean absolute percentage error in percentage, PPMC= Pearson product-moment correlation coefficient. $(*)$ significance level of $= 0.05$. (a) Parameter based on ground truth data.

the case of the strawberry field, crop density was 0.36 plants/m^2.

The forecast events for counting maize and strawberry plants present positive MBE values or a tendency to over-count crops in the plot (Table 3). In the maize field, MAE's magnitude is almost four plants per plot. In the strawberry field, the MAE value is more than one plant in a plot. The Ae value is 4.1 and 7.8 respectively for the maize and strawberry.

4 Discussion

4.1 Effect of the monitoring system and the system resolution for plant counting

A relevant task is the development of approaches which can be generalized across measuring systems, different fields and/or different crops. By using three different sensor and UAV setups, this contribution confirms that the previously tested and used pipeline (Lottes et al., 2018a), can be generalized to evaluate RGB UAV-based images. Nevertheless, the principal difference was observed during the establishment of a flight plan, setup 1 in comparison with setup 2 and 3 (Table 1), presented a bigger sensor size (4/3-inch against 1-inch) and a longer focal length (45 against 8.8 mm), that allows a greater field of view in high resolution as mentioned in Pepe et al. (2018), and therefore this system was able to map bigger field area with the same flight time. No advantages in terms of reduced flight lines and use of fewer GCPs were observed for the setup 2 with real-time kinematic (RTK) system (Rabah et al., 2018), probably related to the small size of the mapped area and the less challenging and planar characteristics of the observed object, in this case, the trial field with small vegetation during first growing stages.

The principal changing factor, the ground sample distance, is not a

limiting factor for the dataset with lower resolution (GSD of 0.3 cm). Sa et al. (2018) mentioned that a resolution of 0.85 - 1.18 cm (GSD) can compromise performance of weed detection by using a Deep Convolutional encoder-encoder architecture for crop/weed segmentation. This study found that in early season a classification between crops and weeds is difficult because of morphological similarities. Using image resolutions of 0.02, 0.2, and 0.5 cm (GSD) and a similar image processing approach, Lottes et al. (2017) mentioned that field arrangement is the best supporter for detection of crop and weeds, highlighting the importance for row detection and the spatial relationships among multiple individual plants which remain constant at the mentioned image resolutions.

4.2 Sky conditions

In the past, it was affirmed that the use of spectral images in high resolution for the analysis of vegetation could negatively affect the classification of areas under shadows. This phenomenon could lead to a false interpretation of physiological and metabolic activities (Zarco-Tejada et al., 2013). In our time-series case using RGB-images, the lowest forecast error is acquired under clear sky conditions, which tells us that the used FCN pipeline approach could be used in shaded scenarios without losing performance. Another possibility is to expect a lower MAPE value under more favorable sky conditions; this means that under the same date and place of experimental field, but different sky condition like cloudy/ no-shaded scenario, the forecast error can be lower than 3.96% (Table 3). The employed time series analysis has to be contrasted with similar categories of crop growth stage conditions (BBCH 12, 12-14 and 14-16) and weed pressure and in various sky conditions to determine the effect of shadow in the counting performance.

4.3 Intra-row distance as a critical factor for a FCN pipeline

In this study, the experimental field with 11-12 cm intra-row distance presented a high emergence of wheat from previous seasons, correctly classified as weeds. We furthermore avoid using herbicide in this experiment to increase the chances of emerging intra-row weeds. The detection of sugar-beet stems and hence counting of crop plants can be performed with an average precision of 78.8% by using UAV-RGB images in grass plants' presence according to past results (Lottes et al., 2018a), supporting observed outputs of the time-series analysis. In the past, a similar study based on mask R-CNN approach using robot-based RGB images for detection of common beans and maize presented similar precision values for crop segmentation (60-80%) and highlighted the importance of the weed cover for misdetection (Champ et al., 2020), this study found that small weed cover presents a high probability to be pixel-wise classified and counted as a crop by calculating the barycenter. A YOLOv3 CNN architecture also mentioned a high mean average precision of 86% for counting stems in cotton seedling, although the effect of weed pressure was not evaluated, there are some disadvantages by overlapping and detecting small seedling (Oh et al., 2020).

The intra-row distance of 21-22 cm presented, over all categories of growing stages, less counting error than the intra-row distance of 11-12 cm. In the past, CNN and FCN pipelines for crop/weed detection used images containing crops arranged in a typical distribution, 15-25 cm of intra-row distance, and 30-60 cm distance between rows (Sa et al., 2018; Lottes et al., 2018b). For the employed pipeline, images for training presented crops distributed in the same fashion. By observing the output images, an explanation of the higher rate of under-counting crops by low

intra-row distance is that neighbor sugar-beet stems are difficult to count. Sometimes two plants are counted as one, mostly if the midrib and tip leaf is aligned and substantially close to the neighbor crop stem (Figure 6a).

4.4 Importance of growing stage for successful counting

In this paper, one of our main purpose is to specify optimal conditions for plant counting. In the vegetation period, crop growth is dynamic, and UAV-flights to capture images have to be performed on time to obtain the best results. To provide crop numbers, the determination of stem position in different growth stages is a crucial point, the performance to detect stems of plants with two-leaf to later growth stages can achieve 95.5% of precision and 98.0% of recall according to a past experience, but the best growth stage was not specified (Lottes et al., 2019). This study specifies that the optimal growth stage is dependent on intra-row distance. For the case of practice fields or intra-row distance of 20-21 cm and 45 cm between rows, the flight has to be performed in preference between 16-18 BBCH stage compared with the BBCH 14-16. In experimental fields with an intra-row distance of 11-12 cm, UAV-flight provides the best results with the BBCH stage of 14-16.

4.5 RGB-UAV for monitoring sugar beet/weed competition

In a past experiment, Lottes et al. (2018b) determined plant-soil cover by labeling pixel-wise images into crop and weed; for their objective, they used RGB-NIR images containing sugar beets in different growth

stages obtained from field robot platform and a similar FCN pipeline. After optimization and analysis of performance, the pixel-wise approach achieved 91% of F1-score. Our study is based on the above-mentioned excellent performance to determine the soil cover of weeds and sugar beets. In this way, we have used a similar approach that focuses on UAVs' importance for agronomic application. This includes using the advantage of UAVs for flexible image acquisition in the agronomic goal of weed control (Peña et al., 2013) and the possibility to support the decision for weeding during the entire critical period of competition, which usually takes place between BBCH 14 and 32 during approximately 21-28 days under normal growing conditions (Petersen, 2004). A point to clarify in future studies is which threshold of sugar beet/weed coverage is the most effective to reduce yield losses by using a specific weed control method.

4.6 Pipeline extension to additional crops

By analyzing the pipeline outputs, maize and strawberries' over-counting is related to crop counting of stems belonging to weeds located in the intra-row space (Figure 6b). A similar phenomenon of overestimation was observed in the past, but not by over-counting but by over-segmenting (Bosilj et al., 2019). The mentioned work also transferred learning from a SegNet CNN model by retraining from sugar beets to onions data. This tendency of overestimation must slightly impact the Precision from the transferred crop and is an aspect to study in future investigations. Despite this, calculated results show low counting errors, and as mentioned on Sec. 4.1 crops distribution, and regular intra-row distance should be the principal reasons for controlling the counting after retraining a model on a small scale.

In this extension, the counting error for other dicot as strawberry is higher than the error from a monocot as maize. The same effect was

observed in a past study using transfer of knowledge of a CNN approach from sugar beets to carrots and onions (Bosilj et al., 2019). In this experience, the monocot crop presented 2% less performance drop than a dicot crop by a pixel-wise classification of the input image. Quan et al. (2019) mentioned, that a pixel-wise segmentation of maize seedlings can achieve a segmentation of up to 98% using RGB-images and a faster R-CNN approach, and sunny conditions can negatively influence the performance of seedling detection. Furthermore, a precision of up to 95% was reported using RetinaNet and CenterNet CNN architectures for counting maize seedling (Karami et al., 2020). Nevertheless, the performance of this approach could not be contrasted due to the lack of a labeled testing dataset. This paper confirms the feasibility of using a pipeline developed and modeled for sugar beets and applying it with less labeling effort on maize and strawberry fields. Counting error is lower than 4% and less or as high as the experimental field with low intra-row distance.

5 Conclusion

Monitoring and counting crops on the field are of high interest for farmers, experimental fields, and the seed-producing industry. The presented UAV-RGB image processing pipeline can deliver the number of sugar beets on the field with an error lower than 4.6%. The intra-row distance and the growing stage of sugar beets are relevant parameters for accurate plant counting. This evaluated variables present the most effective constellation with a crop distance of 21-22 cm and at BBCH 16-18. In experimental-field plant density, with an intra-row distance of 11-12 cm, the smallest forecast error for crop number is shown at BBCH 12-14. The extension of the previously trained FCN pipeline to other crops is possible with a small training dataset, the errors of predictions are lower than 4% by evaluating

practical fields of maize and strawberry, which highlights the potential use of the image processing approach to wide numbers of crops. Overall, considering the parameters as mentioned earlier, automatic monitoring of crop fields using UAV-images followed by proper processing of these can output reliable information that increases efficiency in the crop production by reducing the manual counting effort of the farmers.

6 Author contributions

conceptualization, A.B, P.L., C.S., S.B., A.K.M and S.P.; methodology, P.L., A.B., F.I., C.S. and S.P.; validation, A.B, P.L., N.A.W and S.B.; formal analysis, P.L., A.B., F.I and S.P.; investigation, A.B, P.L., N.A.W and S.B.; resources, P.L., C.S, S.B. and A.K.M; writing–original draft preparation, A.B. and P.L.; writing–review and editing, all authors; visualization, A.B., P.L. and S.P.; supervision, C.S., A.K.M. and S.P; funding acquisition, P.L., C.S. and A.K.M.

7 Acknowledgments

We would like to thank Sebastian Streit and Dennis Grunwald for proofreading and their helpful comments and suggestions regarding the structure. We thank Dirk Koops and Jonathan Eggers for supporting the ground truth labelling and measurement. The authors are grateful with Judith Berger from ARGENORD for delivering information crucial in the introduction. The authors thank the breeders KWS (Einbeck, Germany),BetaSeed GmbH (Frankfurt, Germany), MariboHilleshög GmbH (Hannover, Germany), SesvanderHave GmbH (Eisingen, Germany) and Strube D&S GmbH (Söllingen, Germany); for the provision of seeds in field trials.

8 Funding:

This study was partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2070 - 390732324. Part of research staff was partially founded within Coordination Beet Research International (COBRI) and the Farmerspace project which is supported by funds of the Federal Ministry of Food and Agriculture (BMEL) based on a decision of the Parliament of the Federal Republic of Germany. The Federal Office for Agriculture and Food (BLE) provides coordinating support for digitalization in agriculture as funding organisation, grant number FZK 28DE104A18.

9 Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2015). Segnet: a deep convolutional encoder-decoder architecture for image segmentation. *Computing Research Repository (CoRR)*, abs/1511.00561.

Bosilj, P., Aptoula, E., Duckett, T., and Cielniak, G. (2019). Transfer learning between crop types for semantic segmentation of crops versus weeds in precision agriculture. *Journal of Field Robotics*, 37(1):7–19.

Champ, J., Mora-Fallas, A., Goëau, H., Mata-Montero, E., Bonnet, P., and Joly, A. (2020). Instance segmentation for the fine detection of

628 crop and weed plants by precision agricultural robots. *Applications in*
629 *Plant Sciences*, 8(7).

630 Cioni, F. and Maines, G. (2010). Weed control in sugarbeet. *Sugar Tech*,
631 12(3):243–255.

632 Deng, L., Mao, Z., Li, X., Hu, Z., Duan, F., and Yan, Y. (2018). UAV-
633 based multispectral remote sensing for precision agriculture: A com-
634 parison between different cameras. *ISPRS Journal of Photogrammetry*
635 *and Remote Sensing*, 146:124–136.

636 Dumoulin, V. and Visin, F. (2016). A guide to convolution arithmetic for
637 deep learning. *ArXiv*, abs/1603.07285.

638 Durrant, M., Brown, S., and Bould, A. (1985). The assessment of
639 the quality of sugar-beet seed. *The Journal of Agricultural Science*,
640 104(1):71–84.

641 Karami, A., Crawford, M., and Delp, E. J. (2020). Automatic plant
642 counting and location based on a few-shot learning technique. *IEEE*
643 *Journal of Selected Topics in Applied Earth Observations and Remote*
644 *Sensing*, 13:5872–5886.

645 Kato, T. (2016). Prediction of photovoltaic power generation output and
646 network operation. In *Integration of Distributed Energy Resources in*
647 *Power Systems*, pages 77–108. Elsevier.

648 Kunz, C., Schröllkamp, C., Koch, H.-j., Eßer, C., and Lammers, P. S.
649 (2015). Potentials of post-emergent mechanical weed control in sugar
650 beet to reduce herbicide inputs. *Landtechnik*, 70(3):67–81.

651 Lottes, P., Behley, J., Chebrolu, N., Milioto, A., and Stachniss, C.
652 (2018a). Joint stem detection and crop-weed classification for plant-

specific treatment in precision farming. *IEEE International Conference on Intelligent Robots and Systems*, pages 8233–8238.

Lottes, P., Behley, J., Chebrolu, N., Milioto, A., and Stachniss, C. (2019). Robust joint stem detection and crop-weed classification using image sequences for plant-specific treatment in precision farming. *Journal of Field Robotics*, 37(1):20–34.

Lottes, P., Behley, J., Milioto, A., and Stachniss, C. (2018b). Fully convolutional networks with sequential information for robust crop and weed detection in precision farming. *IEEE Robotics and Automation Letters*, 3(4):2870–2877.

Lottes, P., Khanna, R., Pfeifer, J., Siegwart, R., and Stachniss, C. (2017). UAV-based crop and weed classification for smart farming. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3024–3031.

Lu, H., Liu, L., Li, Y. N., Zhao, X. M., Wang, X. Q., and Cao, Z. G. (2021). TasselNetV3: Explainable plant counting with guided upsampling and background suppression. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–15.

Märländer, B. (1990). Einfluss der bestandesdichte auf ertrags- und qualitätskriterien sowie über mögliche ursachen der konkurrenz in zuckerübenbestanden. *Agronomy & Crop Science*, 130.

McCool, C., Perez, T., and Upcroft, B. (2017). Mixtures of lightweight deep convolutional neural networks: Applied to agricultural robotics. *IEEE Robotics and Automation Letters*, 2(3):1344–1351.

Milošević, M., Vujaković, M., and Karagić, D. (2010). Vigour tests as indicators of seed viability. *Genetika*, 42(1).

679 Oh, S., Chang, A., Ashapure, A., Jung, J., Dube, N., Maeda, M., Gonza-
680 lez, D., and Landivar, J. (2020). Plant counting of cotton from UAS
681 imagery using deep learning-based object detection framework. *Remote*
682 *Sensing*, 12(18):2981.

683 Peña, J. M., Torres-Sánchez, J., de Castro, A. I., Kelly, M., and López-
684 Granados, F. (2013). Weed mapping in early-season maize fields using
685 object-based analysis of unmanned aerial vehicle (UAV) images. *PLoS*
686 *ONE*, 8(10):e77151.

687 Pepe, M., Fregonese, L., and Scaioni, M. (2018). Planning airborne
688 photogrammetry and remote-sensing missions with modern platforms
689 and sensors. *European Journal of Remote Sensing*, 51(1):412–436.

690 Petersen, J. (2004). A review on weed control in sugarbeet. In *Weed*
691 *Biology and Management*, pages 467–483. Springer Netherlands.

692 Pospišil, M., Pospišil, A., and Rastija, M. (2000). Effect of plant density
693 and nitrogen rates upon the leaf area of seed sugar beet on seed yield
694 and quality. *European Journal of Agronomy*, 12(1):69–78.

695 Quan, L., Feng, H., Lv, Y., Wang, Q., Zhang, C., Liu, J., and Yuan,
696 Z. (2019). Maize seedling detection under different growth stages
697 and complex field environments based on an improved faster R–CNN.
698 *Biosystems Engineering*, 184:1–23.

699 Rabah, M., Basiouny, E.-G., Ghanem, E., and Elhadary, A. (2018). Using
700 RTK and VRS in direct geo-referencing of the UAV imagery. *NRIAG*
701 *Journal of Astronomy and Geophysics*, 7(2):220–226.

702 Rahman, M. and Wang, Y. (2016). Optimizing intersection-over-union
703 in deep neural networks for image segmentation. volume 10072, pages
704 234–244.

Sa, I., Popović, M., Khanna, R., Chen, Z., Lottes, P., Liebisch, F., Nieto, J., Stachniss, C., Walter, A., and Siegwart, R. (2018). WeedMap: a large-scale semantic weed mapping dramework using aerial multispectral imaging and deep neural network for precision farming. *Remote Sensing*, 10(9):1423.

Schreiner, A. J., Unger, D. A., Menzel, W. P., Ellrod, G. P., Strabala, K. I., and Pellet, J. L. (1993). A comparison of ground and satellite observations of cloud cover. *Bulletin of the American Meteorological Society*, 74(10):1851 – 1862.

Shcherbakov, M., Brebels, A., Shcherbakova, N., Tyukov, A., Janovsky, T., and Kamaev, V. (2013). A survey of forecast error measures. *World Applied Sciences Journal*, 24:171–176.

Smit, A. (1993). The influence of sowing date and plant density on the decision to resow sugar beet. *Field Crops Research*, 34(2):159–173.

Söğüt, T. and Arioglu, H. (2004). Plant density and sowing date effects on sugarbeet yield and quality. *Journal of Agronomy*, 3(3):215–218.

Tong, P., Han, P., Li, S., Li, N., Bu, S., Li, Q., and Li, K. (2021). Counting trees with point-wise supervised segmentation network. *Engineering Applications of Artificial Intelligence*, 100(February).

Wu, X., Aravecchia, S., Lottes, P., Stachniss, C., and Pradalier, C. (2020). Robotic weed control using automated weed and crop classification. *J Field Robotics*.

Zarco-Tejada, P., Guillén-Climent, M., Hernández-Clemente, R., Catalina, Á., González, M.-R., and Martín, P. (2013). Estimating leaf carotenoid content in vineyards using high resolution hyperspectral im-

agery acquired from an unmanned aerial vehicle (UAV). *Agricultural and Forest Meteorology*, 171-172:281–294.

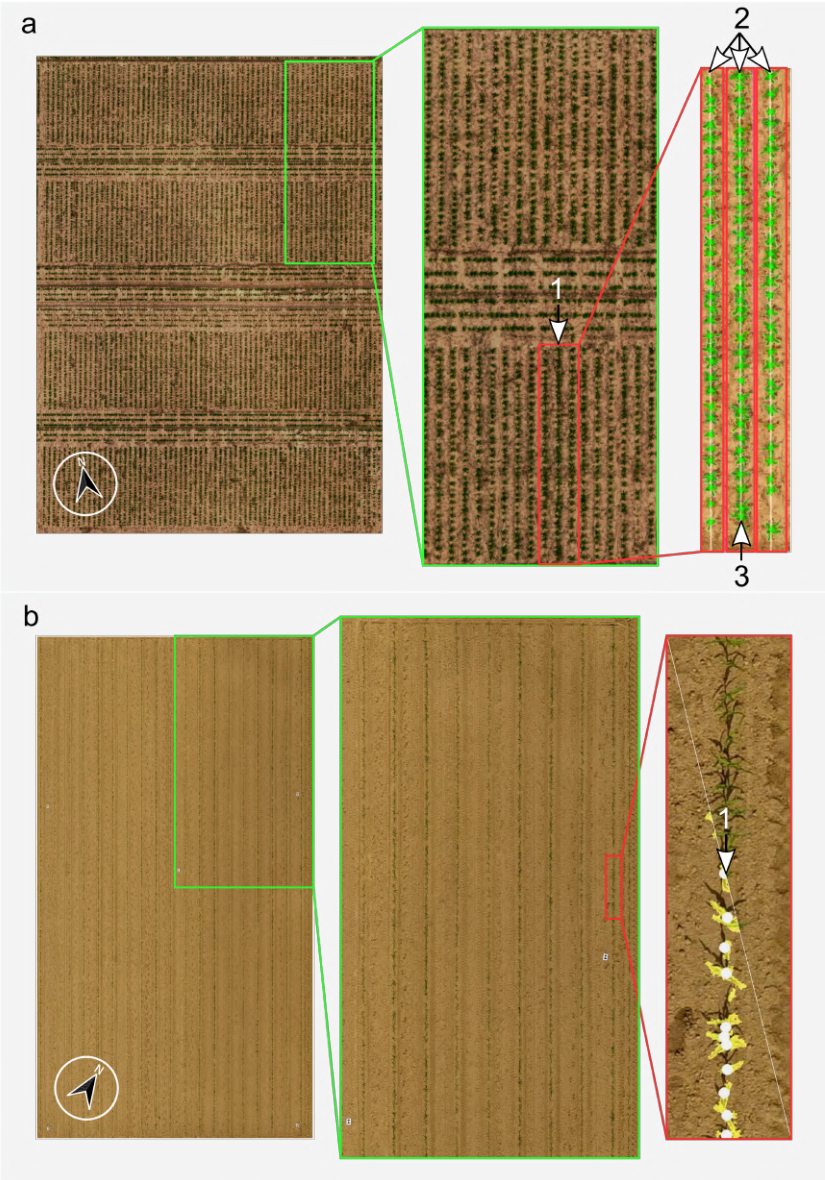

Figure 1: **Experimental fields**: (a) UAV-acquisition and analysis of a sugar-beet field at 28th May 2019 in Börßum (location 2), 1) plot segmentation, 2) row detection, 3) stem detection and crop counting. (b) UAV-acquisition and analysis of maize field in Klein-Altendorf (location 4), 1) stem detection and plant counting

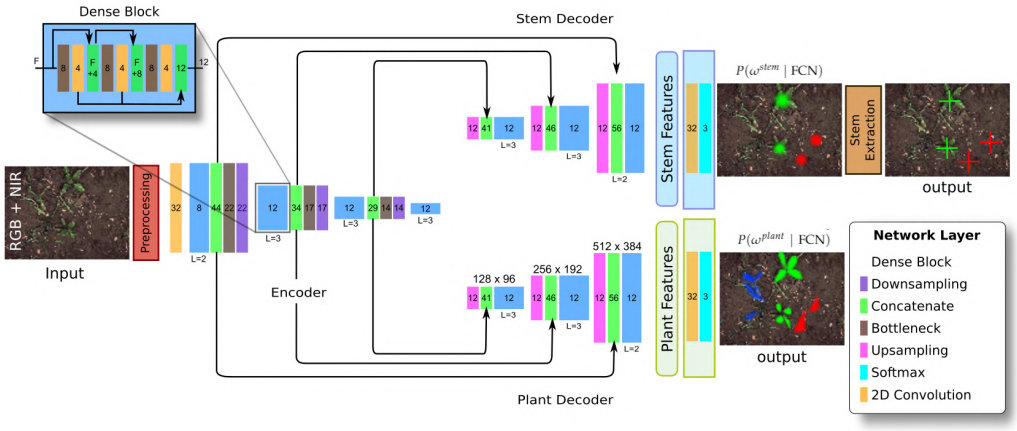

Figure 2: **FCN architecture**. Encode of input images using the encoder and following by the pass the feature volumes to the task-specific decoders, the stem and the plant decoder. Obtained outputs were: the plant mask considering the classes crop, weed, soil for the pixel-wise classification of the plants, and the stem mask considering the classes crop, weed, soil for the segmentation crop-weed stem regions. Finally, the extraction of the stem positions from the stem mask in the stem extraction. Inside the layers, it is shown the number of output features maps. L represents the number of stacked consecutive 2D convolutional layers.

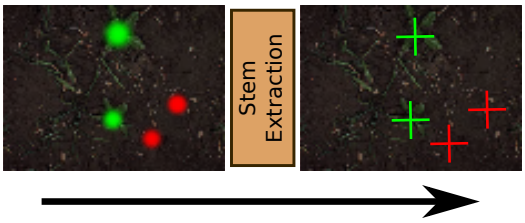

Figure 3: **Extraction of pixel-wise stem locations** by computing a weighted center of mass of the stem regions predicted by the FCN. For the weighting, it was considered the predicted probabilities for each pixel belonging to a stem region.

Figure 4: Illustration of **difficult conditions for counting plants** using vision-based classification approaches. Left: Mutually overlapping sugar beets. Right: Due to narrow seeding, the sugar beets overlap early after the emergence phase. In addition, individual and contiguous plants are separated by straw in the image space.

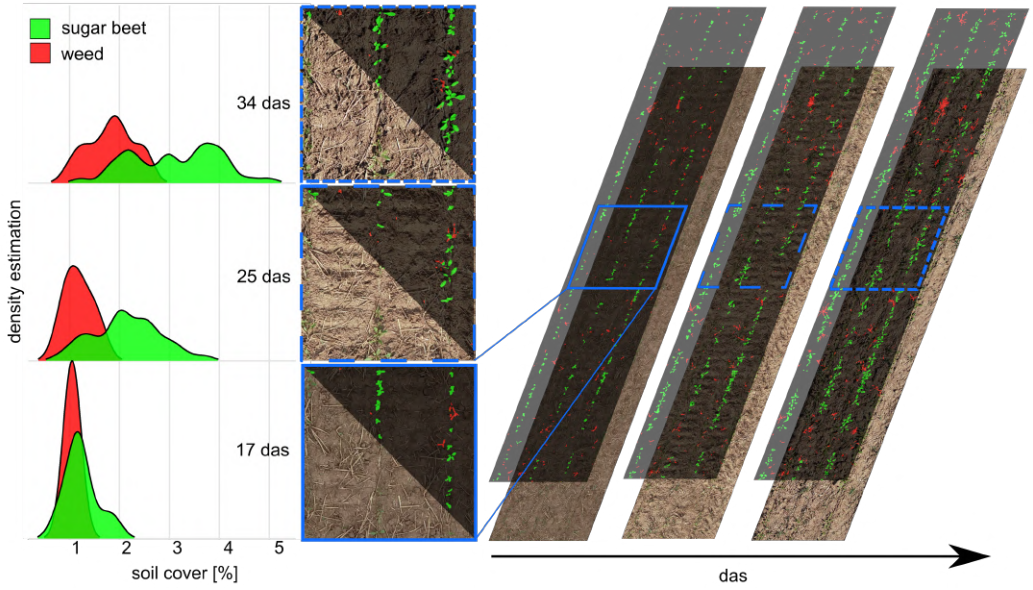

Figure 5: **Development of vegetation cover**: time series labeling of sugar beet (green) and weed (red) covered area at 17, 25 and 34 das. (*) Density estimation of 40 plots of trial field located in Höckelheim (location 3). Evaluated field area 9.5 m^2.

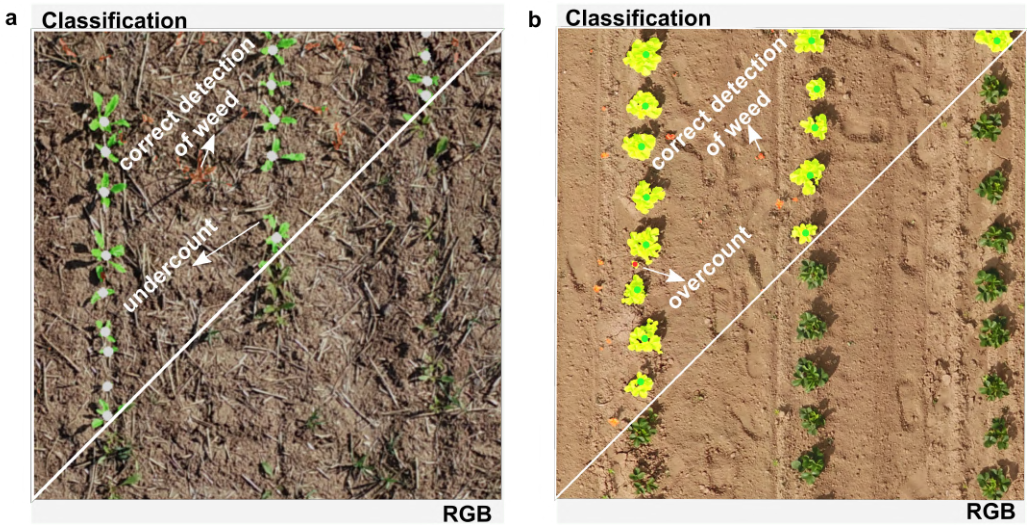

Figure 6: **Counting performance** after classification in the upper left triangle: (a) counted sugar-beet plants (white dots) and one under-count of sugar-beet stem at 34 days after sowing in Höckelheim - location 3 (b) performance of a sugar-beet image processing pipeline extended to a strawberry field, counted strawberry (green dots) over-count case of false classification of weed (red dot) - location 5. The lower right triangle shows the original RGB patch image.