Computers and Electronics in Agriculture 202 (2022) 107388

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Deep learning-based early weed segmentation using motion blurred UAV images of sorghum fields

Nikita Genze [a,b,*], Raymond Ajekwe [d], Zeynep Güreli [a,b], Florian Haselbeck [a,b], Michael Grieb [d], Dominik G. Grimm [a,b,c,*]

[a] Technical University of Munich, TUM Campus Straubing for Biotechnology and Sustainability, Bioinformatics, Schulgasse 22, 94315 Straubing, Germany
[b] Weihenstephan-Triesdorf University of Applied Sciences, Bioinformatics, Petersgasse 18, 94315 Straubing, Germany
[c] Technical University of Munich, Department of Informatics, Boltzmannstr. 3, 85748 Garching, Germany
[d] Technology and Support Centre in the Centre of Excellence for Renewable Resources (TFZ), Schulgasse 18, 94315 Straubing, Germany

ARTICLE INFO

Keywords:
Deep learning
Weed detection
Weed segmentation
UAV
Precision agriculture

ABSTRACT

Weeds are undesired plants in agricultural fields that affect crop yield and quality by competing for nutrients, water, sunlight and space. For centuries, farmers have used several strategies and resources to remove weeds. The use of herbicide is still the most common control strategy. To reduce the amount of herbicide and impact caused by uniform spraying, site-specific weed management (SSWM) through variable rate herbicide application and mechanical weed control have long been recommended. To implement such precise strategies, accurate detection and classification of weeds in crop fields is a crucial first step. Due to the phenotypic similarity between some weeds and crops as well as changing weather conditions, it is challenging to design an automated system for general weed detection. For efficiency, unmanned aerial vehicles (UAV) are commonly used for image capturing. However, high wind pressure and different drone settings have a severe effect on the capturing quality, what potentially results in degraded images, e.g., due to motion blur. In this paper, we investigate the generalization capabilities of Deep Learning methods for early weed detection in sorghum fields under such challenging capturing conditions. For this purpose, we developed weed segmentation models using three different state-of-the-art Deep Learning architectures in combination with residual neural networks as feature extractors.

We further publish a manually annotated and expert-curated UAV imagery dataset for weed detection in sorghum fields under challenging conditions. Our results show that our trained models generalize well regarding the detection of weeds, even for degraded captures due to motion blur. An UNet-like architecture with a ResNet-34 feature extractor achieved an F1-score of over 89% on a hold-out test-set. Further analysis indicate that the trained model performed well in predicting the general plant shape, while most misclassifications appeared at borders of the plants. Beyond that, our approach can detect intra-row weeds without additional information as well as partly occluded plants in contrast to existing research.

All data, including the newly generated and annotated UAV imagery dataset, and code is publicly available on *GitHub*: https://github.com/grimmlab/UAVWeedSegmentation and Mendeley Data: https://doi.org/10.17632/4hh45vkp38.4.

Abbreviations: SSWM, Site Specific Weed Management; ML, Machine Learning; DL, Deep Learning; SVM, Support Vector Machine; RF, Random Forest; CNN, Convolutional Neural Network; FCN, Fully Convolutional Network; GSD, Ground Sampling Distance; OBIA, Object Based Image Analysis; NIR, Near Infra-Red; GIMP, GNU Image Manipulation Program; TL, Transfer Learning; CRF, Conditional Random Field; TPE, Tree Parzen Estimator; DS, Dice Score or Sorensen-Dice Coefficient; TP, True Positives; FP, False Positives; FN, False Negatives; ASPP, Atrous Spatial Pyramid Pooling; CLAHE, Contrast Limited Adaptive Histogram Equalization.

* Corresponding authors at: Technical University of Munich, TUM Campus Straubing for Biotechnology and Sustainability, Bioinformatics, Schulgasse 22, 94315 Straubing, Germany.

E-mail addresses: nikita.genze@tum.de (N. Genze), dominik.grimm@hswt.de (D.G. Grimm).

https://doi.org/10.1016/j.compag.2022.107388
Received 18 February 2022; Received in revised form 31 August 2022; Accepted 8 September 2022
Available online 18 September 2022

1. Introduction

Weeds are undesired plants in agricultural fields that affect crop yield and quality by competing for nutrients, water, sunlight and space (Patel and Kumbhar, 2016). For centuries, farmers have used several strategies and resources to remove weeds. Some of these strategies include mechanical (mowing or tilling) and chemical (herbicides) control. The use of herbicides remains the most common control strategy as it is less time consuming and labor-intensive (Aktar et al., 2009). However, recent studies show that this strategy can impact negatively on the biota and the surrounding environment over time (Holt, 2004; Who, 1990). These disadvantages lead to a social pressure requesting less usage (Ridgway et al., 1978). To reduce the amount of herbicide and impact caused by uniform spraying, site-specific weed management (SSWM) through variable rate herbicide application and mechanical weed control using an autonomous field robotic system have long been recommended (Fernández-Quintanilla et al., 2018). These approaches heavily depend on weed sensing and mapping technology. To achieve this, an accurate as well as efficient detection and classification of weeds in crop fields is a crucial first step (Liu and Bruch, 2020).

Owing to the phenotypic similarity between some weeds and crops as well as the dynamic weather conditions, designing an automated system for general weed detection in crop fields is challenging (Hasan et al., 2021). Furthermore, variability in the emergence time of different weeds at various crop growth stages makes it rather complicated to design a generalized model. Most studies in weed detection have leveraged plant vegetation indices (Hamuda et al., 2016; Meyer and Neto, 2008) and hand-engineered features with a focus on classical Machine Learning (ML) methods, such as Support Vector Machines (SVM) and Random Forests (RF) (Zheng et al., 2017). For example, Kazmi *et al.* employed conventional image processing based on vegetation indices for creeping thistle weed detection in sugar beet fields (Kazmi et al., 2015). Relatedly, Islam *et al.* reported weed detection in chili plants using a similar technique (Islam et al., 2021).

Deep Learning (DL), particularly Convolutional Neural Networks (CNNs), are capable of extracting relevant features from raw images and were used extensively in image classification in recent years (LeCun and Bengio, 1998). They generalize well across illuminations and obstructions and were applied for several agricultural tasks (Genze et al., 2020; Kussul et al., 2017; Grinblat et al., 2016; Chen et al., 2017). These properties make them well suited for weed classification using drone images from fields, which comprise several changing conditions, such as illumination. Furthermore, recent surveys show a superiority of DL-based approaches compared to classical ML (Liu and Bruch, 2020; Kamilaris and Prenafeta-Boldu, 2018; Kamilaris et al., 2017; Alom et al., 2019). Additionally, several authors have concluded that hand engineered features tend to be less precise and hard to generalize compared to DL-based approaches (Penatti et al., 2015; Huang et al., 2018).

One downside of CNNs is the amount of data needed for generalization (Sun et al., 2017). There are several publicly available datasets for weed detection, as summarized in a recent review (Lu and Young, 2020). However, most of these datasets were collected using high-resolution cameras mounted on field robots or pulled carts. DL models have been applied for the task of weed detection on different crops. For example, Ramirez *et al.* used a dataset called WeedMap (Sa et al., 2018), consisting of high-resolution orthomosaic images of a sugar beet field in Germany. The authors compared SegNet (Badrinarayanan et al., 2015) and a modified U-Net (Ronneberger et al., 2015) with a recent segmentation architecture called DeepLabv3 (Chen et al., 2017). DeepLabv3 led to a higher classification accuracy due to a greater spatial context (Ramirez et al., 2020). Huang *et al.* studied weeds in rice fields and showed that Fully Convolutional Networks (FCNs) (Long et al., 2015) outperform other methods in accuracy and efficiency for drone imagery with a ground sampling distance (short: GSD, distance between adjacent pixel centers) of 0.3 cm (Huang et al., 2018; Huang et al., 2018).

Regarding data generation, most weed detection studies rely on object-based image analysis (OBIA) (Lam et al., 2021; Hay and Castilla, 2006), unsupervised (Dian Bah et al., 2018; dos Santos Ferreira et al., 2017) or semi-supervised (Pérez-Ortiz et al., 2015; Lottes and Stachniss, 2017) approaches for semantic segmentation. The main reason for this is the large effort of a manual pixel-based labeling. Beyond that, additional information like the spatial distribution of the crop rows (Lottes and Stachniss, 2017; Onyango and Marchant, 2003; Milioto et al., 2018) or different sensors (e.g. near infrared (NIR) and multispectral) are widely used (Lottes et al., 2018). By using different light spectra, it is easier to separate vegetation from soil (Yeom et al., 2019), thus making the ground truth generation faster. One downside of current NIR and multispectral sensors is their low GSD compared to RGB sensors. Although NIR and multispectral sensors provide more information about the vegetation compared to RGB sensors, the GSD is one of the most critical parameters for accurate weed detection. Besides the difficulty of detecting small weeds captured by drones with low GSD, it is more time-consuming and error-prone to generate the ground truth.

For a real-world application, an efficient way for image capturing is needed, for which drones can be utilized (Mukherjee et al., 2019). On the one hand, a higher throughput and temporal resolution are the main advantages compared to ground robots. On the other hand, image capturing using drones is prone to motion blur, e.g. due to wind-related displacements. This is an additional challenge for a robust weed classification. Only a limited number of datasets have been generated using drone captures, due to the time-consuming generation and labeling of such datasets.

In this study, we investigate weed segmentation capabilities of DL-based approaches using real-world motion blurred drone captures in sorghum fields. For this purpose, we generated a manually curated and expert-labeled UAV dataset of early weeds and sorghum plants. Our dataset is fully and manually annotated with the highest precision possible without relying on OBIA or threshold-based segmentation techniques. This dataset enabled us to evaluate state-of-the-art DL architectures in combination with residual neural networks of different sizes as feature extractors for early weed and sorghum segmentation. We further, evaluated the generalization abilities of our trained models with respect to different growth stages of sorghum.

2. Material and methods

In the next section, we provide insights into the data acquisition and annotation. Afterwards, we outline the machine learning models, the hyperparameter optimization as well as all used evaluation metrics. Finally, we describe the experimental setup.

2.1. Data acquisition and annotation

Images collected by an UAV on an experimental sorghum field in Southern Germany provided data for this study. The sorghum variety "Farmsugro 180" was sown at 37.5 cm row spacing with a seeding density of 25 seeds per m^2. During image capture, several weed species were present on the field. These weeds comprised mostly of dicotyledons namely, Goosefoot (*Chenopodium album* L.), Field pennycress (*Thlaspi arvense*), Wild chamomile (*Matricaria chamomilla*), Common gypsyweed (*Veronica officinalis*) and Cotton thistle (*Onopordum acanthium*). Beyond that, a consumer-grade drone "DJI Mavic 2 Pro" fitted with a 20 MP Hasselblad camera (L1D-20c) that captures images with a resolution of 5472x3648 $pixels^2$ was used. Automated drone flight with camera pointing nadir and a capture overlap of ten percent was carried out at a flight altitude of five meters above ground level. At this altitude, the corresponding GSD was one millimeter, and therefore precise enough to recognize sorghum and weeds in early growth stages. A set of 60 high resolution images was captured in late spring when sorghum was at the growth stage 17 (according to the BBCH scale (Hess et al., 1997). Thereby, a time- and battery-efficient setting was selected with a non-

stop capturing at 60 points of equal distance along the flight path, which leads to motion blur in the collected images.

In this work, we defined the weed detection task as distinguishing weeds from sorghum plants and soil pixel by pixel (weed segmentation). To enable supervised ML, we assigned one of these three classes (weed, sorghum, soil) to every pixel of each collected image. For the annotation process, we used the open-source GNU Image Manipulation Program (GIMP). Agronomy experts guided the labeling process to ensure a high-quality ground truth. Fig. 1 shows an example of a drone capture and its related pixel-wise ground truth.

About 6.300 patches from 19 images were annotated resulting in over 3.000 sorghum and 3.000 weed instances, as summarized in Table 1. We divided each image in non-overlapping patches with a resolution of 256x256 pixel for an efficient training process. Additionally, all non-square patches (image and labels) were padded with zeros.

Several factors influenced the degree of motion blur in the images, e.g. the current velocity of the drone or external factors such as wind. The blurriness of images varied between different captures, e.g. some captures were slightly sharper after the rotation of the drone at an edge of the flight plan, because the drone was flying slower and stabilized itself after the rotation. A detailed evaluation of the blurriness of our captures can be found in Supplementary Text 1. Another challenge in this real-world datasets are overlapping plants. This might lead to multiple plants being labeled and counted as one instance.

To evaluate the generalization capabilities of our model, we labeled two additional UAV missions, as shown in Table1. These images were captured during different UAV flights on another part of the field and show sorghum at the growth stages BBCH 15 and 19. More detailed information and statistics about the dataset can be found in the Supplementary Figure S1.

Due to the pixel-wise annotation of the datasets, it can be easily adapted to other tasks, such as patch-based classification or object detection using bounding boxes. These additional labels can be generated automatically using our fine-grained labeled segmentation masks.

2.2. Machine learning and experimental setup

The architecture of our weed segmentation model consists of a feature extractor and a semantic segmentation architecture.

For our weed segmentation model, we combined these six semantic segmentation architectures with four residual neural networks of different sizes as feature extractors (He et al., 2016; Veit et al., 2016). Residual networks combat the vanishing gradient problem by introducing identity skip connections and allowing data to flow from any layer directly to any subsequent layer. This enabled researchers to build deeper networks with hundreds of layers efficiently. In addition, they utilize convolutions with strides instead of pooling layers for downsampling. The feature extractors ResNet-18, 34, 50 and 101 were used in this study, which consist of 18, 34, 50 and 101 layers, respectively.

Table 1
Summary statistics of the datasets used in this study.

Dataset Name	Sorghum BBCH	Annotated Patches	Sorghum Instances	Weed Instances
sorghum_17	17	6270	3056	3060
sorghum_15	15	110	115	429
sorghum_19	19	110	99	981

With respect to the semantic segmentation architecture, we investigated six state-of-the-art DL architectures, namely UNet (Ronneberger et al., 2015), DeepLabv3+ (Chen et al., 2017) and four variants of FCN (Long et al., 2015), since these models have already been successfully used for other crops (Ramirez et al., 2020; Huang et al., 2018). The main aspects of the model architectures are illustrated in Fig. 2.

FCNs perform convolutions, pooling and upsampling and can produce a spatial segmentation map of the same size as the input image and scale across different image sizes, as indicated in Fig. 2a-d. Semantic information from deep and coarse layers can be combined with appearance information from shallow and fine layers in the network. The FCN-32s model does not combine information of different layers, but predicts one segmentation mask from the last layer of the network and then calculates the result using bilinear interpolation with a factor of 32 (Fig. 2a). In the FCN-16s model, the predictions of the last two layers are combined. Therefore, the predictions from the last layer must be upsampled by a factor of two first and then combined with the feature maps of the second-last layer. In addition, one-by-one convolutions are used to match the depth of both feature maps before combining them. Finally, the result is upsampled by a factor of 16 (Fig. 2b). In the FCN-8s model, the two last feature maps are concatenated first (as in the FCN-16s model). This result gets upsampled by a factor of two again to combine it with the third-last layer. Finally, this output can only be upsampled by eight to match the spatial resolution of the input image (Fig. 2c). The FCN-8s model is able to predict more fine-grained segmentation maps by combining predictions from three different layers. This helps in training a model that is able to predict finer details. Alternatively, dilated (atrous) convolutions (Fisher and Koltun, 2016) can be used to preserve the spatial resolution. In addition, they use a greater receptive field to cover more information in the layer. This led to the architecture FCN-8s+ dilation, as shown in Fig. 2d. This architecture is similar to FCN-32s, as no information from intermediate layers is used, but it only needs to interpolate the predictions bilinear with a factor of eight.

The UNet architecture was initially developed for biomedical tasks and has been adapted to different domains. Existing research shows that

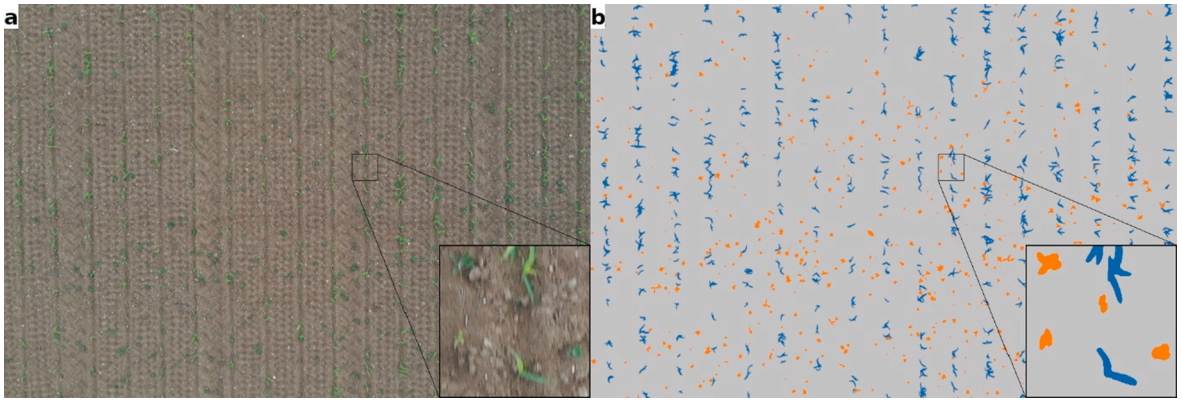

Fig. 1. Example drone capture and related ground truth for the dataset *sorghum_17*. a UAV capture from one of our sorghum fields **b** Pixel-wise ground truth with background in gray, sorghum plants in blue and weed plants in orange.

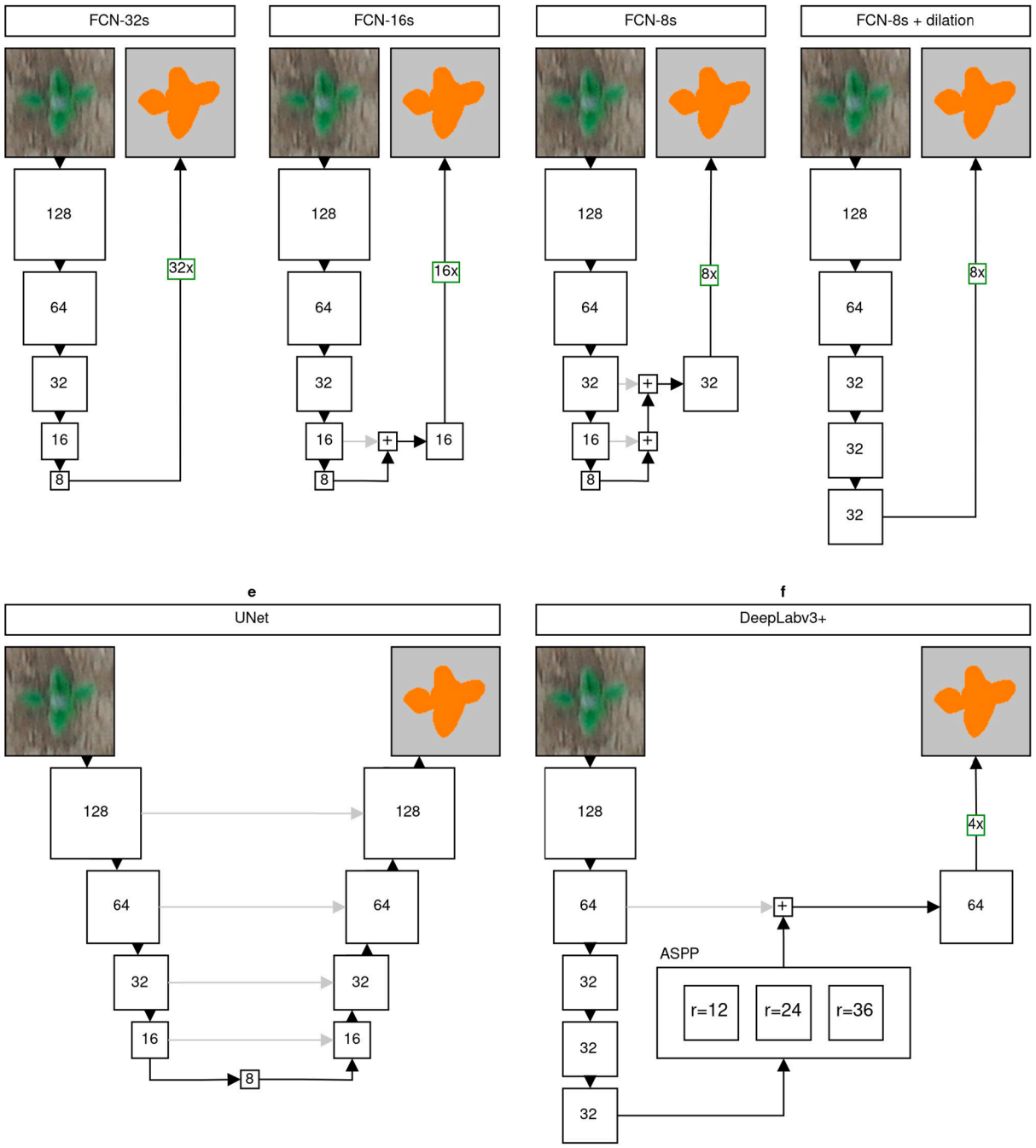

Fig. 2. Model diagrams of the different semantic segmentation architectures used in this study. Numbers indicate the spatial resolution of the feature maps. Skip connections are shown as gray arrows. The concatenation of feature maps is shown as plus. Upsampling operations before the concatenation is omitted for readability. Final bilinear interpolation is shown in a green box indicating the upsampling factor. **a-c** Modifications of the FCN architecture with different strides. **d** FCN architecture using dilated convolutions, preserving the spatial resolution in the last layers. **e** UNet architecture using skip connections. **f** DeepLabv3+ architecture using Atrous Spatial Pyramid Pooling (ASPP) with atrous rates of r = 12, 24 and 36.

this model achieved good segmentation results even with little training data by relying on a vigorous data augmentation pipeline. Skip connections between encoder and decoder are used to fuse information of all layers of the encoder with the corresponding layers in the decoder (Fig. 2e). In addition, the feature maps are upsampled more gradually in the decoder compared to FCN. Finally, the upsampling is done using learnable filters instead of a fixed bilinear interpolation.

DeepLabv3+ is a model architecture presented by Chen *et al.* that improves the coarse spatial resolution - caused by repeated pooling layers - by using dilated (atrous) convolutions. These special operations adjust the field-of-view of the filter, as indicated in Fig. 2f. The Atrous Spatial Pyramid Pooling (ASPP) module uses multiple parallel-dilated convolutions with different rates to consider objects at different scales (Chen et al., 2017). Similar to FCN and UNet, an encoder-decoder architecture is used to recover the lost spatial information and refine the predictions at the object boundaries.

Furthermore, early stopping (Prechelt et al., 2012) and a learning rate scheduler were implemented to lower the computational cost and avoid overfitting. The Adam (Kingma and Ba, 2014) optimizer was used with different initial learning rates. Following the principles of risk minimization (Vapnik, 1992), a differentiable version of the Sørensen-Dice Coefficient (DS) was used as a loss function to update the weights in

the training process. The same loss was applied to validate the performance of the trained models. This metric is commonly used to measure the similarity of two samples in the evaluation of semantic segmentation tasks (Bertels et al., 2019) and has been shown to be robust to class imbalance (Sudre et al., 2017). It is defined as follows:

$$DS = \frac{2*TP}{2*TP + FP + FN} \quad (1)$$

where TP is the number of true positives, FP is the number of false positives and FN is the number of false negatives. The batch size was fixed to 100 for each combination of feature extractor and segmentation architecture.

Hyperparameters can greatly influence the predictions of a DL-based model. There are several strategies for optimizing these, e.g. Grid or Random Search (Bergstra and Bengio, 2012). Contrary to these strategies, Tree-structured Parzen Estimator (TPE) (Bergstra and Bengio, 2012) selects a new set of hyperparameters based on previous experiments conducted in an optimization study. Using an objective function, a new set of hyperparameters is sampled to achieve a high probability of a good score. This strategy requires the construction of a probabilistic model for the optimization process, but is efficient due to the automatic selection of promising settings. For this reason, we employed this strategy to optimize the hyperparameters summarized in Table 2.

All models were trained and evaluated based on patches of 256x256 pixel^2. We excluded patches that contained only background or small plant instances (>99% only background pixels). This resulted in 2156 patches from 12 UAV captures. Patches that were zero-padded in an earlier step made up 143 patches (=6.63%). Padding does not affect the classification accuracy (Hashemi, 2019) itself, but is needed to use the whole dataset more efficiently. The data was split into four folds for cross validation and an additional hold-out test-set for estimating the final model performance, as shown in Fig. 3. The averaged performance on the validation sets was used to determine the best performing hyperparameter combinations for each weed segmentation model. A total of 50 different hyperparameter sets were evaluated. The setup leading to the smallest error averaged over all validation sets was trained again on the complete training and validation data. Finally, we evaluated the performance of the resulting model on the hold-out test-set.

We report the evaluation metrics precision, recall and the F1-score on a pixel basis:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

Table 2
Summary of the hyperparameters used and optimized in this study.

Hyperparameter	optimized	Range	Description
initial learning rate	yes	10^{-4} – 10^{-2}	learning rate at the start of model training. It is sampled from a log-uniform distribution.
learning rate decay	yes	0.9–0.1	rate of the learning rate decay used when the loss function plateaus. The step size was set to 0.1.
learning rate scheduler patience	no	5	reduce the learning rate, if the validation loss did not improve over this number of epochs
batch size	no	100	Number of patches that can be sent through the network in one forward pass
early stopping patience	no	10	stop the training loop, if the validation loss did not improve over this number of epochs

$$F1 - score = \frac{Precision*Recall}{Precision + Recall} \quad (4)$$

Further, as we observed a high-class imbalance, we calculate the macro-averaged metrics, as weight-averaged metrics favor the majority class (background). Weight-averaged metrics consider the proportion for each label in the dataset. As the weed class is occurring the least, but is the most important one, macro-averaged metrics are more suitable, as they reflect the arithmetic mean of all classes.

The final performance was calculated after combining all patches into individual images. This procedure is independent of the batch size and the size of the patches, thus ensures the comparison to different other weed detection approaches in future studies.

Beyond that, data augmentation (Shorten and Khoshgoftaar, 2019) is a common technique in image processing to enlarge the amount of training data. It is widely used, as the performance of a ML-based model depends on the amount of available data. We used several ways to augment data, which are summarized in Table 3. In addition to the standard translational augmentations, we used Contrast Limited Adaptive Histogram Equalization (CLAHE) (Zuiderveld, 1994) to enhance the local contrast of the image patches together with a channel-wise normalization.

We trained our weed segmentation models using Transfer Learning (TL) (He et al., 2016). We initialized the feature extractors using weights pre-trained on ImageNet (Deng et al., (2009–2009)). The decoder was initialized using He initialization as proposed by He et al. (He et al., (2015–2015)).

All code is implemented in Python 3.8 using the packages numpy (Mukherjee et al., 2019); pytorch (Hess et al., 1997); torchvision (Marcel and Rodriguez, 2010); albumentations (He et al., 2016); optuna (Veit et al., 2016); kornia (Fisher and Koltun, 2016) and scikit-learn (Prechelt et al., 2012). Optimizations were conducted using an Ubuntu 20.04 LTS machine with 104 CPU Cores, 756 GB memory and four GeForce RTX 3090 GPUs. Each model was trained on a single GPU. Models that did not fit on this GPU were trained on another machine using a NVIDIA A40 GPU with 48 GB VRAM.

3. Results

In this section, we first provide an overview of our results. Then, we analyze the predictions of our best performing weed segmentation model in more details. Finally, we discuss our results.

3.1. Results overview

First, we evaluated the architecture choice of the segmentation model in combination with four ResNet feature extractors employing a fourfold cross-validation. In Table 4, we show the results on the validation sets for these experiments using the Sørensen-Dice Coefficient (DS) summarized over all 50 trials, see Section 2.2. In summary, UNet with ResNet-34 as feature extractor performed best.

In general, the deepest feature extractor ResNet-101 performed worse for all segmentation models. The best performing feature extractor was dependent on the selected segmentation architecture. We observed minor differences in the objective value when comparing different feature extractors while using the same semantic segmentation model. In addition, there were minor differences in the objective values between different segmentation models UNet and DeepLabv3+. When using dilated convolutions in the feature extractor with the FCN-8s architecture (no concatenation of intermediate feature maps), the objective values had only minor differences compared to UNet and DeepLabv3+. This indicates a minor effect of the selection of the semantic segmentation model.

Both, FCN-16s and FCN-8s combined multiple output layers together and then upsampled the result. Here we observed a higher objective value in general, meaning worse predictions. In addition, the standard

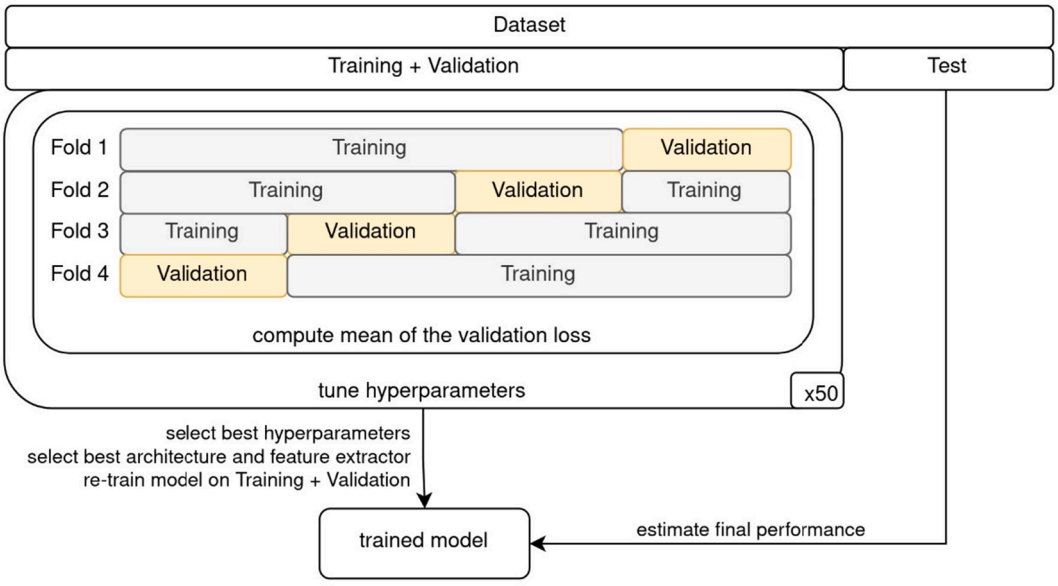

Fig. 3. Overview of the model selection and evaluation process conducted for each weed segmentation model.

Table 3
Data augmentation techniques used in this study.

Augmentation Transform	Description
Horizontal flip	Flips the image horizontally
Vertical flip	Flips the image vertically
RandomRotate90	Randomly rotates the image by a factor of 90 degrees
Transpose	Swaps rows and columns of the image
CLAHE	Applies CLAHE to the input image

Table 4
Best evaluation results on the validation sets for each weed segmentation model based on the Sørensen-Dice Coefficient for 50 trials. Standard deviation is provided in brackets. The best performing feature extractor for each segmentation architecture is denoted in bold.

	ResNet-18	ResNet-34	ResNet-50	ResNet-101
FCN-32s	0.0281 (0.0061)	0.0285 (0.0061)	**0.0275 (0.0056)**	0.0278 (0.0059)
FCN-16s	**0.0226 (0.16)**	0.0232 (0.16)	0.0255 (0.16)	0.0270 (0.16)
FCN-8s	**0.0134 (0.16)**	0.0136 (0.17)	0.0159 (0.17)	0.0160 (0.18)
FCN-8s + dilation	0.00999 (0.0095)	0.0101 (0.0044)	**0.00951 (0.0062)**	0.00963 (0.0052)
UNet	0.00923 (0.0041)	**0.00910 (0.0022)**	0.00913 (0.0033)	0.00926 (0.0035)
DeepLabv3+	0.0118 (0.0018)	0.00945 (0.0022)	**0.00939 (0.0022)**	N.A.*

* More than 48 GB of GPU memory were required.

deviation between the trials is approximately 28 times higher for FCN-16s and FCN-8s compared to FCN-32s. This holds true for all feature extractors examined in this study.

The best combination UNet with ResNet-34 was more robust during hyperparameter optimization, indicated by a small standard deviation of 0.0022.

3.2. Best performing model

Furthermore, the final performance was estimated on the above-described hold-out test-set using the best performing weed segmentation model (UNet with ResNet-34), which we re-trained on the complete training and validation data using the hyperparameter values that led to the lowest error. The final model has high macro-averaged evaluation values on the hold-out test-set, as shown in Table 5. Every metric is above 86% indicating sufficient generalization capabilities.

Fig. 4a shows the model's accuracy more precisely on a per-class basis by using a normalized confusion matrix. The soil was almost correctly predicted with high accuracy of 99.9%. The model classified 86.1% of all sorghum pixels correctly while misclassified 11.3% as background and 2.6% as weed. More importantly, 72.7% of all weed pixels were classified correctly, while predicting 23.2% as background and only 4.1% as sorghum. Additional per-class metrics can be found in Supplementary Table S5.

In addition, we tested the generalization abilities of our model (trained on sorghum at BBCH 17) using two different growth stages of sorghum (BBCH 15 and 19). Most importantly, our weed segmentation model trained on BBCH 17 is still able to detect weeds at least as well on images with lower or higher growth stages of sorghum (see Fig. 4b and 4c). Interestingly, for sorghum at BBCH 19, 77.6% of all weed pixels were classified correctly (i.e. an increase of 4.8% compared to results shown in Fig. 4a). However, the predictions on sorghum for different growth stages dropped to 50.0% for BBCH 15 (a decrease of 36.9%), mostly predicting weeds correctly. However, for BBCH 19 there was a decrease of only 1.1% to 85.8%, indicating consistent high performance on larger BBCH stages of sorghum.

Beyond that, we performed a qualitative analysis of the results on the hold-out test set and the additional patches from different growth stages. In general, the predictions of our weed segmentation model are accurate for the hold-out test set (as shown in Fig. 5a-f).

Most plants were segmented correctly, capturing the general shape of the plants. However, we observed most confusions at the borders of the plants. Unfortunately, these misclassifications of mainly the borders cannot be captured by the used evaluation metrics. Furthermore, most pixels belonging to "old weeds" (i.e. larger weeds that have been already present when sowing the crop) are predicted correctly, as shown in Fig. 5b, i and k. Only some parts of these instances were mis-classified as

Table 5
Macro-averaged results on the hold-out test-set in percent for the best performing weed segmentation model with UNet and ResNet-34.

DS	Precision	Recall	F1-score
99.69	93.01	86.25	89.37

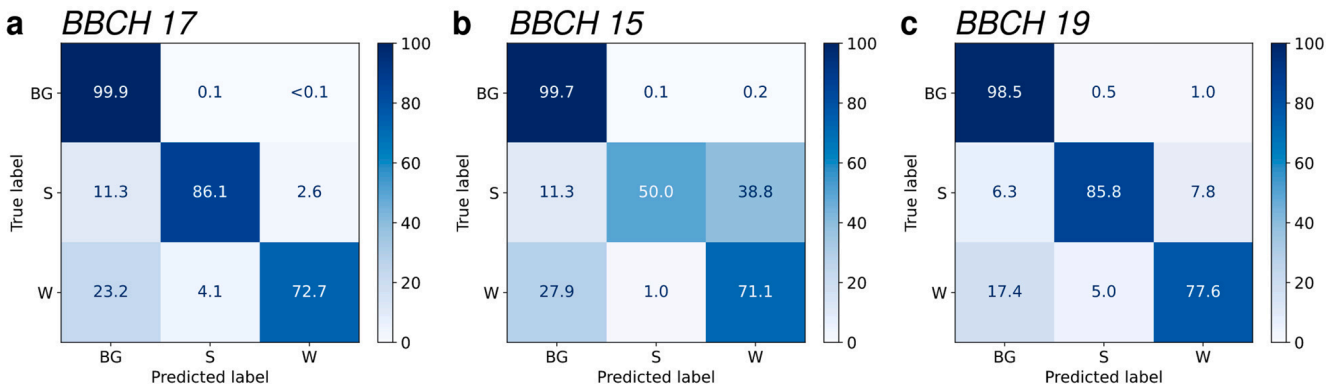

Fig. 4. Normalized confusion matrix by support size in percent shows the pixel-based classification results on the hold-out test-set. Background/soil is denoted by BG, sorghum by S and weed by W. **a** Sorghum at growth stage BBCH 17**. b** Sorghum at growth stage BBCH 15**. c** Sorghum at growth stage BBCH 19**.**

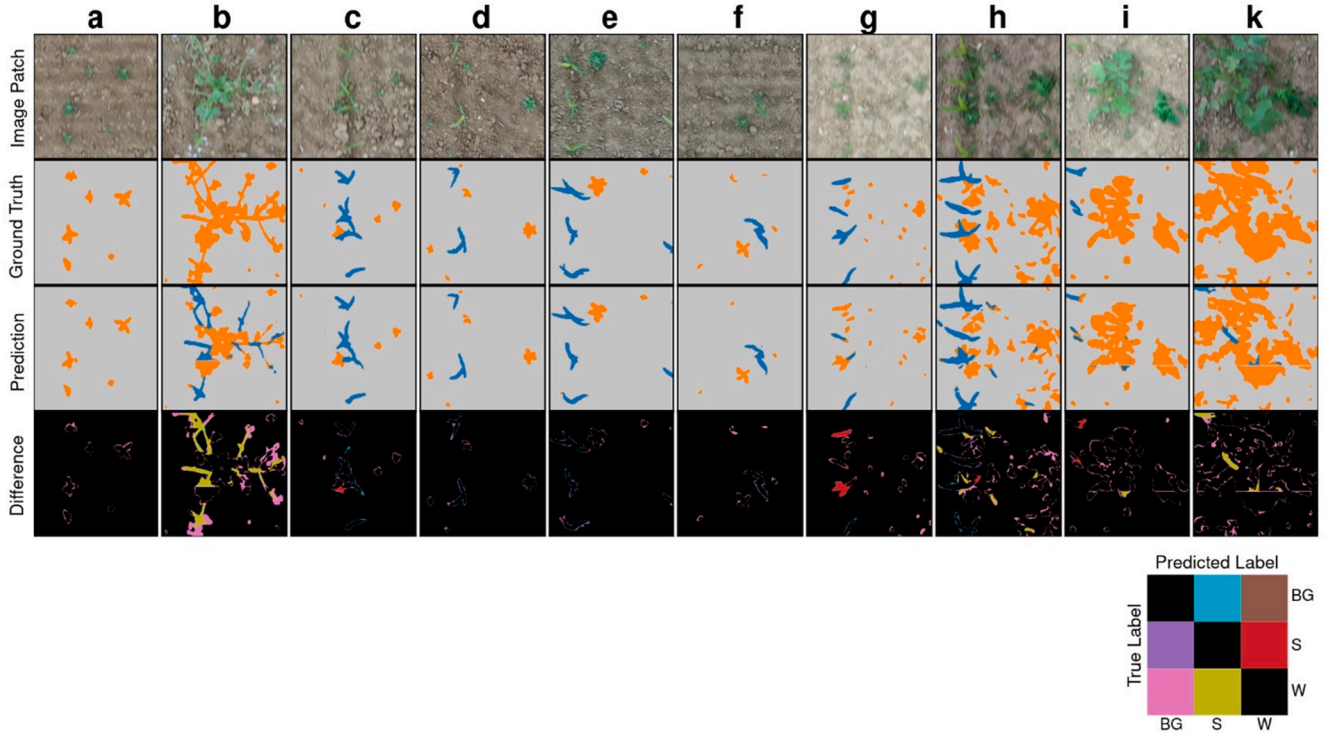

Fig. 5. Qualitative results on the hold-out test-set. Image patches of size 400x400 pixel^2 are cropped from each test image to show more details. Background (BG) is colored in gray, sorghum (S) in blue and weed (W) pixels in orange. The difference map shows only the misclassifications between ground truth and prediction.**a-c** Examples with a high degree of motion blur. **a** The general shape of weeds is predicted correctly. **b** Large weed plants that could not be removed before sowing the field. Most pixels are predicted correctly. There are small artifacts visible, which are due to the patching process**. c** Weeds intersecting with sorghum plants were predicted correctly**. d-f** Examples with a low degree of motion blur showing weeds and sorghum plants from different captures. The general shape is predicted correctly, where **d** is a patch from test_04, **e** from patch test_05 and **f** from patch test_06. **g-k** Comparison of two additional drone flights with different growth stages and different illumination. **g** Weed infestation when sorghum was at BBCH15, showing leaves of sorghum being predicted as weed. **h** The same excerpt as in c when sorghum was at BBCH 19. **i** Large weed plants that could not be removed before sowing the field is predicted mostly correct. Sorghum was at BBCH 15. **k** The same patch when sorghum was at BBCH 19.

sorghum, even though not many "old weed" instances were in the training set. Additionally, intra-row weeds were correctly detected, even if they are in close proximity or overlapping the crop (see Fig. 5c).

Plants in patches with a higher relative sharpness (for the sharpness calculations see Supplementary Text 1) were also predicted correctly, as shown in Supplementary Figure S5 and Fig. 5d-f. This indicates a high generalization ability between different degrees of motion blur.

For the additional growth stages, the main drawback of this model was the prediction of sorghum plants as weeds for BBCH 15 (Fig. 5g and 5i). This might be because of morphological changes due to growth (i.e. additional leaves were visible in BBCH 17 compared to BBCH 15). Here, in most of the examples only a part of the plant was misclassified, but occasionally there were also complete sorghum plants predicted as weed. The general weed distribution and the shape of large weeds was predicted correctly. Only some parts of weeds were predicted as sorghum. For BBCH19, the model was still able to predict weeds correctly, despite the fact, that they had grown larger than the weeds the model was trained on. In addition, "old weeds" that grew even larger were mostly predicted correctly, as shown in Fig. 5k. Some artifacts are visible especially for larger plants, which are due to the patching process, as one plant was partly cut in different patches resulting in small parts present on the patch border. These small parts were then not predicted correctly,

as denoted in yellow in Fig. 5i and k.

Finally, the images of BBCH 15 and BBCH 19 were captured with different illumination settings than the model was trained on. Interestingly, weeds in these captures were still predicted with high precision. This indicates that our model might be applicable to captures of different image quality.

The complete dataset and code are publicly available on Mendeley-Data[1] and in our *GitHub* repository.[2]

3.3. Discussion

In this paper, we present a DL-based approach to weed detection and segmentation in sorghum fields under real-world conditions using motion blurred UAV images. Our dataset was generated with a high GSD of one millimeter by conducting drone missions from a low altitude (five meters) in sorghum fields. This might be contrary to several other approaches, which employed a higher flight altitude to ensure higher throughput. Yet, this high GSD allowed us to label and predict smaller weeds potentially resulting in a more accurate weed detection model. In addition, the ground truth was generated using a precise approach with a pixel-based semantic segmentation and expert-curated annotations for the complete dataset. There are several color- or threshold-based approaches, as discussed in a recent survey (Hamuda et al., 2016). The quality of those segmentations might vary considerably with the imaging and weather conditions, potentially resulting in a wrong ground truth, especially when using drone captures degraded by motion blur. As mentioned in this survey, changing weather and imaging conditions should not influence DL-based approaches, as one could use data augmentation to adjust the exposure and saturation.

Due to wind-related displacements, motion blur is a common challenge in drone imagery of agricultural landscapes. However, the effect of real motion blur on weed detection models is not studied widely. In this work, we show that DL models are able to accurately segment weeds from crop in blurry drone images independent on the degree of motion blur. The main bottleneck when developing new supervised DL models is the amount of data to label. We show that sorghum plants were segmented with a high precision in general when using around 3.000 labeled instances. One reason for this might be the size and the shape of the sorghum plants in comparison with weed instances. The crop plants were larger and had a smaller variance in shape. Additionally, overlapping sorghum plants were predicted sufficiently. Sharp edges were not visible due to motion blur and therefore the model could not detect any edges to predict the correct class. It had to focus on different features of the plants to discriminate between sorghum and weed. The model learned the background class well and was able to generalize on different backgrounds. Our dataset had few instances of “old weeds” (large weeds already present when sowing the crop) that were outliers from the normal weed’s distribution. Although the shape was not predicted correctly, still most of these weed pixels were correct. Our model did not segment small weeds, as this is also a general bottleneck in DL approaches and implies that conducting UAV missions with a higher GSD is favorable. However, the overall spatial weed distribution still matches the ground truth, which is the main objective in SSWM. More important, our model is not dependent on the detection of crop rows and is able to detect intra-row weeds, as shown in Fig. 5c and d. Consequently, crops that were not sown in precise rows could be detected using our model.

In this study, UNet with ResNet-34 performed best. This indicates that a large feature extractor (ResNet-101) is overfitting towards the training data more easily. There were slight differences in performance when comparing the architectures of the segmentation models. UNet was only slightly better than DeepLabv3+ and FCN-8s with dilated convolutions, indicating that the choice of the semantic segmentation model is not important when using UAV captures with a high GSD of 0.1 cm. The design choice of using dilated convolutions in the FCN-8s model achieved the highest relative improvement. The UNet architecture yielded the best results, indicating that a more gradual upsampling with skip connections is more favorable than atrous convolutions used in DeepLabv3+.

In our study, we observed a discrepancy between evaluation metrics and a qualitative assessment of the actual predictions. The confusion matrix treats every pixel equally, which might not be useful in weed segmentation approaches. By further analyzing the model’s predictions, we could conclude that most mis-classified pixels were on the border of plants. However, with a real-world SSWM application in mind, plant borders are not that important, as the general shape is sufficient to generate accurate weed maps. A correct classification of crop and weed pixels is more important, as this is the main challenge. Furthermore, rare occurring but large weeds and the high percentage of the background class distorted the results obtained from the confusion matrix, as misclassification of those weeds had a higher impact. In our dataset, 98% of all pixels were background, which had a significant impact on per-pixel metrics and the confusion matrix.

The main goal of weed detection for SSWM is to generate a weed density map, so weeds can be removed in downstream tasks. These tasks (i.e., weed removal by a field robot or herbicide application) are not pixel-precise, so arguably a coarser ground truth might be also sufficient while lowering the annotation cost and time. Our dataset could act as a base for further weed research using drones, as a coarser ground truth (i.e. for object detection or patch-based classification) could be generated automatically. One major difficulty in using a pixel-based segmentation approach is the ambiguity of the ground truth labels due to a subjectivity in the labeling process. This ambiguity is challenging for model training and evaluation especially when using real-world motion blurred imagery.

The application of our model on other BBCH stages showed good generalization capabilities towards the larger growth stage (BBCH 19). When presented with imagery of a smaller growth stage (BBCH 15), our model wrongly predicted around 39% of sorghum pixels as weed. Further analysis showed that a majority of the misclassifications were present on smaller sorghum plants where only two or three leaves were visible. This indicates that our model needs to be re-trained with UAV imagery of different growth stages in order to be more robust and applicable for SSWM.

Some errors in the predictions were also due to the patch generation process. They happen only, when there is a small part of a plant present on a patch border. The network was trained on mostly intact plants, so when presented with a partly cut plant, it might misclassify it. In addition, background pixels are mostly predicted in close proximity of the patch borders. These errors are not severe, as they only occur for larger plants that were separated in the patching process and might be solved by using a sliding window approach and generating patches with overlap.

While our model performed sufficiently well on our hold-out test-set, there are several factors that were not evaluated in this study. Our test-set was captured on a certain agricultural field. However, the weed flora might be different in other regions or fields. Thus, it is likely that there are several other weed species growing on different agricultural landscapes. In addition, the sampling time (time of the capture) of the different datasets is similar, as we conducted the drone missions at noon. Different sampling times (morning, noon or evening) and different weather conditions (sunny, cloudy) might have an effect on the illumination (white balance, color temperature) of the resulting captures, which might have a negative effect on the model’s performance. Nevertheless, our experiments on the datasets sorghum_15 and sorghum_19 with different illuminations indicate good generalization of our model. In this study, we briefly investigate the generalization capabilities of our model on two different growth stages. Our results

[1] https://doi.org/10.17632/4hh45vkp38.4.
[2] https://github.com/grimmlab/UAVWeedSegmentation.

suggest that it might be necessary to fine-tuning our model on different growth states which would improve the overall performance. A larger dataset with additional growth stages could be beneficial to yield a unified model for early weed detection across different stages. Additionally, drone settings, i.e. flight speed and altitude might play an important role on the quality of the captures and might influence the predictions as well. Therefore, we plan to generate a more diverse dataset to evaluate different model architectures on a wider scope, dealing with a more diverse weed flora, weather conditions and drone settings.

4. Conclusion

In this study we present a machine learning framework enabling early weed and sorghum detection and segmentation in real-world, degraded drone images. The model has high segmentation abilities and can detect intra-row weeds as wells as overlapping plants in captures with different degrees of motion blur. Further, a qualitative analysis of the actual predictions indicates that the general form of most plants can be detected accurately, with minor problems at borders of the plants. As discussed, plant borders are not that important in a downstream application for site specific weed management.

Additionally, we show that our method can be applied on different BBCH stages of sorghum. However, the predictions on smaller sorghum plants in an early growth stage dropped and might lead to an overestimation of weed instances. Nevertheless, our work shows promising results and give first insights that we plan to extend in future research.

With this study, we publish the first dataset for weed detection in sorghum, which can be used as a basis for future research. Additionally, we made the code of our proposed method publicly available to ensure the reproducibility of results and allow the comparison in future studies.

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.

Data availability

The generated and labeled dataset and the code for our proposed machine learning–based model can be found on Mendeley data https://doi.org/10.17632/4hh45vkp38.4 and GitHub: https://github.com/grimmlab/UAVWeedSegmentation.

Acknowledgments

This work was supported by the Weihenstephan-Triesdorf University of Applied Sciences and the Technical University of Munich, Campus Straubing for Biotechnology and Sustainability.

Funding

Funding for the research presented in this paper is provided by the Bavarian State Ministry for Food, Agriculture and Forests within the EWIS project (Funding ID: G2/N/19/13).

Authors' contributions

DGG, NG, MG conceived and designed the study. RA conducted the drone flights and the image acquisition. NG and ZG labeled the dataset. RA and MG guided the labeling process with domain expertise. NG implemented the machine-learning pipeline and conducted all computational experiments with support from FH. NG analyzed the results with help from DGG. NG and DGG wrote the manuscript with contributions from all authors. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication.

Not applicable.

Competing interests.

The authors declare that they have no competing interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compag.2022.107388.

References

Aktar, M.W., Sengupta, D., Chowdhury, A., 2009. Impact of pesticides use in agriculture: their benefits and hazards. Interdiscip Toxicol 2, 1–12. https://doi.org/10.2478/v10102-009-0001-7.

Alom, M.Z., Taha, T.M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M.S., Hasan, M., Van Essen, B.C., Awwal, A.A.S., Asari, V.K., 2019. A state-of-the-art survey on deep learning theory and architectures. Electronics (Switzerland) 8 (3), 292.

Badrinarayanan, V., Kendall, A., Cipolla, R., 2015. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation.

Bergstra, J., Bengio, Y., 2012. Random Search for Hyper-Parameter Optimization. J. Mach. Learn. Res. 13, 281–305.

Bertels, J., Eelbode, T., Berman, M. et al., 2019. Optimizing the Dice Score and Jaccard Index for Medical Image Segmentation: Theory and Practice. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 11765 LNCS:92–100. https://doi.org/10.1007/978-3-030-32245-8_11.

Chen, L.-C., Papandreou, G., Schroff, F. et al., 2017. Rethinking Atrous Convolution for Semantic Image Segmentation. 2331–8422.

Chen, S.W., Shivakumar, S.S., Dcunha, S., Das, J., Okon, E., Qu, C., Taylor, C.J., Kumar, V., 2017. Counting Apples and Oranges With Deep Learning: A Data-Driven Approach. IEEE Rob. Autom. Lett. 2 (2), 781–788.

Deng, J., Dong, W., Socher, R., et al., 2009. ImageNet: A large-scale hierarchical image database. In: In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 248–255.

Dian Bah, M., Hafiane, A., Canals, R., 2018. Deep learning with unsupervised data labeling for weed detection in line crops in UAV images. Remote Sens. 10, 1–22. https://doi.org/10.3390/rs10111690.

dos Santos Ferreira, A., Matte Freitas, D., Gonçalves da Silva, G., Pistori, H., Theophilo Folhes, M., 2017. Weed detection in soybean crops using ConvNets. Comput. Electron. Agric. 143, 314–324.

Fernández-Quintanilla, C., Peña, J.M., Andújar, D., Dorado, J., Ribeiro, A., López-Granados, F., Smith, R., 2018. Is the current state of the art of weed monitoring suitable for site-specific weed management in arable crops? Weed Res. 58 (4), 259–272.

Fisher, Yu, Vladlen Koltun, 2016. Multi-Scale Context Aggregation by Dilated Convolutions. In: Yoshua Bengio, Yann LeCun (Eds.). 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings.

Genze, N., Bharti, R., Grieb, M., Schultheiss, S.J., Grimm, D.G., 2020. Accurate machine learning-based germination detection, prediction and quality assessment of three grain crops. Plant Methods 16 (1). https://doi.org/10.1186/s13007-020-00699-x.

Grinblat, G.L., Uzal, L.C., Larese, M.G., Granitto, P.M., 2016. Deep learning for plant identification using vein morphological patterns. Comput. Electron. Agric. 127, 418–424.

Hamuda, E., Glavin, M., Jones, E., 2016. A survey of image processing techniques for plant extraction and segmentation in the field. Comput. Electron. Agric. 125, 184–199.

Hasan, A.S.M.M., Sohel, F., Diepeveen, D., Laga, H., Jones, M.G.K., 2021. A survey of deep learning techniques for weed detection from images. Comput. Electron. Agric. 184, 106067.

Hashemi, M., 2019. Enlarging smaller images before inputting into convolutional neural network: zero-padding vs. interpolation. J. Big Data 6. https://doi.org/10.1186/s40537-019-0263-7.

Hay, A.A.A., Castilla, G., 2006. Object-based image analysis: strengths, weaknesses, opportunities and threats (SWOT). Int. Arch. Photogramm., Remote Sens. Spat. Inform. Sci. 36, 4.

He, K., Zhang, X., Ren, S. et al., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2016-Decem:770–778. https://doi.org/10.1109/CVPR.2016.90.

He, K., Zhang, X., Ren, S., et al., 2015. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In: 2015 IEEE International Conference on Computer Vision (ICCV). IEEE, pp. 1026–1034.

Hess, M., Barralis, G., Bleiholder, H., Buhr, L., Eggers, T.H., Hack, H., Stauss, R., 1997. Use of the extended BBCH scale - General for the descriptions of the growth stages of mono- and dicotyledonous weed species. Weed Res. 37 (6), 433–441.

Holt, J.S., 2004. Principles of Weed Management in Agroecosystems and Wildlands Author (s): Jodie S. Holt Source: Weed Technology, 2004, Vol. 18, Invasive Weed

Symposium (2004), pp. 1559–1562 Published by: Cambridge University Press on behalf of the Weed Science. Weed Technology 18:1559–1562.
Huang, H., Deng, J., Lan, Y., Yang, A., Deng, X., Zhang, L., Gonzalez-Andujar, J.L., 2018. A fully convolutional network for weed mapping of unmanned aerial vehicle (UAV) imagery. PLoS ONE 13 (4), e0196302.
Huang, H., Lan, Y., Deng, J., Yang, A., Deng, X., Zhang, L., Wen, S., 2018. A semantic labeling approach for accurate weed mapping of high resolution UAV imagery. Sensors (Switzerland) 18 (7), 2113.
Islam, N., Rashid, M.M., Wibowo, S., Xu, C.-Y., Morshed, A., Wasimi, S.A., Moore, S., Rahman, S.M., 2021. Early weed detection using image processing and machine learning techniques in an australian chilli farm. Agriculture (Switzerland) 11 (5), 387.
Kamilaris, A., Prenafeta-Boldu, F.X., 2018. Deep learning in agriculture: A survey. Comput. Electron. Agric. 147, 70–90. https://doi.org/10.1016/j.compag.2018.02.016.
Kamilaris, A., Kartakoullis, A., Prenafeta-Boldú, F.X., 2017. A review on the practice of big data analysis in agriculture. Comput. Electron. Agric. 143, 23–37. https://doi.org/10.1016/j.compag.2017.09.037.
Kazmi, W., Garcia-Ruiz, F.J., Nielsen, J., Rasmussen, J., Jørgen Andersen, H., 2015. Detecting creeping thistle in sugar beet fields using vegetation indices. Comput. Electron. Agric. 112, 10–19.
Kingma DP, Ba J (2014) Adam: A Method for Stochastic Optimization.
Kussul, N., Lavreniuk, M., Skakun, S., Shelestov, A., 2017. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. IEEE Geosci Remote Sens. Lett 14 (5), 778–782.
Lam, O.H.Y., Dogotari, M., Prüm, M., Vithlani, H.N., Roers, C., Melville, B., Zimmer, F., Becker, R., 2021. An open source workflow for weed mapping in native grassland using unmanned aerial vehicle: using Rumex obtusifolius as a case study. Eur. J. Remote Sens. 54 (sup1), 71–88.
LeCun, Y.A., Bengio, Y., 1998. Convolutional networks for images, speech, and time series.
Liu, B., Bruch, R., 2020. Weed Detection for Selective Spraying: a Review. Curr. Rob. Reports 1, 19–26. https://doi.org/10.1007/s43154-020-00001-w.
Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: CVPR 2015, pp. 3431–3440.
Lottes, P., Stachniss, C., 2017. Semi-supervised online visual crop and weed classification in precision farming exploiting plant arrangement. In: IEEE International Conference on Intelligent Robots and Systems, 2017-Septe. Institute of Electrical and Electronics Engineers Inc, pp. 5155–5161.
Lottes, P., Behley, J., Milioto, A., Stachniss, C., 2018. Fully convolutional networks with sequential information for robust crop and weed detection in precision farming. IEEE Rob. Autom. Lett. 3 (4), 2870–2877.
Lu, Y., Young, S., 2020. A survey of public datasets for computer vision tasks in precision agriculture. Comput. Electron. Agric. 178, 105760 https://doi.org/10.1016/j.compag.2020.105760.
Marcel, S., Rodriguez, Y., 2010. Torchvision the machine-vision package of torch. In: Del Bimbo, A., Chang, S.-F., Smeulders, A. (Eds.). Proceedings of the international conference on Multimedia - MM '10. ACM Press, New York, New York, USA, pp. 1485.
Meyer, G.E., Neto, J.C., 2008. Verification of color vegetation indices for automated crop imaging applications. Comput. Electron. Agric. 63, 282–293. https://doi.org/10.1016/j.compag.2008.03.009.
Milioto, A., Lottes, P., Stachniss, C., 2018. Real-Time Semantic Segmentation of Crop and Weed for Precision Agriculture Robots Leveraging Background Knowledge in CNNs. In: Proceedings - IEEE International Conference on Robotics and Automation. Institute of Electrical and Electronics Engineers Inc, pp. 2229–2235.
Mukherjee, A., Misra, S., Raghuwanshi, N.S., 2019. A survey of unmanned aerial sensing solutions in precision agriculture. J. Netw. Comput. Appl. 148, 102461 https://doi.org/10.1016/j.jnca.2019.102461.
Onyango, C.M., Marchant, J.A., 2003. Segmentation of row crop plants from weeds using colour and morphology. Comput. Electron. Agric. 39, 141–155. https://doi.org/10.1016/S0168-1699(03)00023-1.
Patel, D.D., Kumbhar, B.A., 2016. Weed and its management: A major threats to crop economy. Journal of. Pharmaceut. Sci. Biosci. Res. 6, 453–758.
Penatti, O.A.B., Nogueira, K., dos Santos, J.A., 2015. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 44–51.
Pérez-Ortiz, M., Peña, J.M., Gutiérrez, P.A., Torres-Sánchez, J., Hervás-Martínez, C., López-Granados, F., 2015. A semi-supervised system for weed mapping in sunflower crops using unmanned aerial vehicles and a crop row detection method. Appl. Soft Comput. 37, 533–544.
Prechelt, L., 2012. Early Stopping — But When? In: Montavon, G., Orr G.B., Müller, K.-R. (Eds.). Neural Networks: Tricks of the Trade, vol. 7700. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 53–67.
Ramirez, W., Achanccaray, P., Mendoza, L.F., et al., 2020. Deep convolutional neural networks for weed detection in agricultural crops using optical aerial images. Int. Arch. Photogramm., Remote Sens. Spat. Inform. Sci. - ISPRS Arch. 42, 551–555. https://doi.org/10.5194/isprs-archives-XLII-3-W12-2020-551-2020.
Ridgway, R.L., Tinney, J.C., MacGregor, J.T., Starler, N.J., 1978. Pesticide use in agriculture. Environ. Health Perspect. 27, 103–112.
Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W.M. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, vol 9351. Springer International Publishing, Cham, pp. 234–241.
Sa, I., Popović, M., Khanna, R., Chen, Z., Lottes, P., Liebisch, F., Nieto, J., Stachniss, C., Walter, A., Siegwart, R., 2018. WeedMap: A large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming. Remote Sens. 10 (9), 1423.
Shorten, C., Khoshgoftaar, T.M., 2019. A survey on Image Data Augmentation for Deep Learning. J. Big Data 6. https://doi.org/10.1186/s40537-019-0197-0.
Sudre, C.H., Li, W., Vercauteren, T., et al., 2017. (2017) Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. Deep Learn. Med. Image Anal. Multimodal Learn. Clin. Decis. Support 2017, 240–248. https://doi.org/10.1007/978-3-319-67558-9_28.
Sun, C., Shrivastava, A., Singh, S. et al., 2017. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. In: ICCV 2017, pp. 843–852.
Vapnik, V., 1992. Principles of risk minimization for learning theory. Adv. Neural Inform. Process. Syst. 831–838.
Veit, A., Wilber, M., Belongie, S., 2016. Residual networks behave like ensembles of relatively shallow networks. Adv. Neural Inform. Process. Syst. 550–558.
Who, 1990. Public health impact of pesticides used in agriculture. World Health Organization, Geneva.
Yeom, J., Jung, J., Chang, A., Ashapure, A., Maeda, M., Maeda, A., Landivar, J., 2019. Comparison of Vegetation Indices Derived from UAV Data for Differentiation of Tillage Effects in Agriculture. Remote Sens. 11 (13), 1548.
Zheng, Y., Zhu, Q., Huang, M., Guo, Y.a., Qin, J., 2017. Maize and weed classification using color indices with support vector data description in outdoor fields. Comput. Electron. Agric. 141, 215–222.
Zuiderveld, K., 1994. Contrast Limited Adaptive Histogram Equalization. In: Graphics Gems. Elsevier, pp. 474–485.