Semantic Segmentation of Crops and Weeds with Probabilistic Modeling and Uncertainty Quantification

Ekin Celikkan[1,2] Mohammadmehdi Saberioon[1] Martin Herold[1] Nadja Klein[3]

[1]GFZ German Research Centre for Geosciences, Potsdam, Germany
{ekin.celikkan, saberioon, herold}@gfz-potsdam.de
[2]Humboldt-Universität zu Berlin, Berlin, Germany
[3]Research Center Trustworthy Data Science and Security, Technische Universität Dortmund, Germany
nadja.klein@tu-dortmund.de

Abstract

We propose a Bayesian approach for semantic segmentation of crops and weeds. Farmers often manage weeds by applying herbicides to the entire field, which has negative environmental and financial impacts. Site-specific weed management (SSWM) considers the variability in the field and localizes the treatment. The prerequisite for automated SSWM is accurate detection of weeds. Moreover, to integrate a method into a real-world setting, the model should be able to make informed decisions to avoid potential mistakes and consequent losses. Existing methods are deterministic and they cannot go beyond assigning a class label to the unseen input based on the data they were trained with. The main idea of our approach is to quantify prediction uncertainty, while making class predictions. Our method achieves competitive performance in an established dataset for weed segmentation. Moreover, through accurate uncertainty quantification, our method is able to detect cases and areas which it is the most uncertain about. This information is beneficial, if not necessary, while making decisions with real-world implications to avoid unwanted consequences. In this work, we show that an end-to-end trainable Bayesian segmentation network can be successfully deployed for the weed segmentation task. In the future it could be integrated into real weeding systems to contribute to better informed decisions and more reliable automated systems.

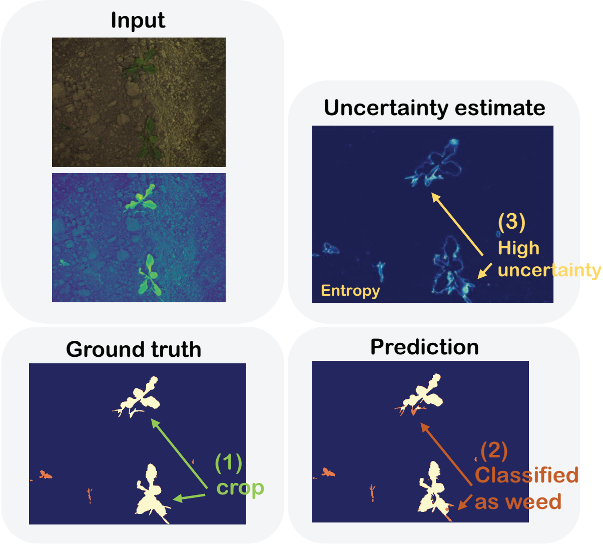


Figure 1: **Probabilistic segmentation of weeds and crops.** Given an input image, our model outputs uncertainty estimates along with semantic segmentation masks. This is useful when, for instance, crop leaves (1) are misclassified as weed (2) by the model. With uncertainty scores, downstream algorithms do not have to rely only on the predicted segmentation masks, but would also consider the (in this case high (3)) prediction uncertainty, and avoid making false decisions (such as spraying the crop).

1. Introduction

Weeds are undesirable plants that tend to overgrow and hinder the growth of desired crops. They are major stressors and cause crop yield loss. Given the ever-growing global demand for food [13], coupled with challenges posed by climate change and environmental degradation [48], any cause of yield loss should be managed and minimized.

Farmers employ various methods to control weed growth in croplands [50]. Chemical or mechanical weeding are common approaches, but come with negative environmental consequences. Excessive herbicide usage can lead to pollution and emergence of herbicide-resistant weeds [35, 33]. Mechanical weeding can inhibit weed regrowth over time [12]. Hence for both methods, if weeding strategies are not implemented efficiently, the financial and environmental burden becomes substantial [21]. Therefore, site-specific weed management (SSWM) is crucial for sustainable agricultural production to meet the increasing demand [1, 16]. SSWM considers the variability within the field, employing tailored weed control methods based on factors such as weed species composition, density, and environmental conditions [16]. Through SSWM, farmers can enhance weed management while minimizing environmental costs.

The biggest building block of SSWM is undoubtedly accurately identifying weed locations. This is a complicated task due to factors like overlapping leaves, occlusion, and similarities between weeds and crops [16]. Dynamic light and weather conditions add further complexity. State-of-the-art segmentation methods are needed to overcome these challenges, and they have already gained prominence in SSWM, offering improved weed detection and classification accuracies [62, 34].

Moreover, given the large size of real life agricultural croplands, it is very difficult for farmers to manually realize SSWM. Automated systems are needed for the efficient management of large-scale fields [4]. And as with any other automated system, to be deployed in a real world application, the segmentation model should not only achieve good overall prediction scores, but ideally it should also know when the prediction is not a confident one. Our probabilistic method combines both of those elements: It predicts accurate segmentation masks and quantifies prediction uncertainty. Therefore it can be easily integrated into existing systems to make informed and accurate decisions, to be used for downstream tasks like automated weeding.

Our contributions can be summarized as follows.

- It is the first end-to-end trainable Bayesian method for weed detection.
- We report uncertainty scores alongside accurate segmentation masks which would improve reliability of corresponding systems.
- Our approach achieves competitive scores in an established dataset for weed segmentation.
- We report results for semantic segmentation of crops and weeds for the first time on the novel PhenoBench [53] dataset (apart from the baseline methods).

By using a probabilistic segmentation approach, we not only obtain more accurate segmentation masks but also acquire uncertainty scores that can be utilized to make informed decisions. We experimentally show that the Bayesian deep learning approach can be used for accurate semantic segmentation of weeds and crops, with the additional benefit of informed confidence on areas of high uncertainty. We hope that our approach would be a step forward to integrating semantic segmentation models into real-life robotic systems for automated weed detection, to be deployed by agricultural experts and farmers.

2. Related Work

2.1. Computer Vision for Weed Phenotyping

The advancements of robotics and computer vision algorithms have triggered efforts for automated quantification and removal of weeds [61, 45, 55, 4]. The existing works for weed monitoring can be roughly grouped into two categories: Object-based and pixel-based methods.

Early work on object-based weed recognition rely mostly on classical machine learning techniques [27, 39]. The authors classify weeds using unmanned aerial vehicle (UAV) images. They first detect vegetation in the field by normalized difference vegetation index (NDVI), then use several statistical and spatial features for classification with random forests [3]. The evaluation is done in per-plant basis and the method relies on the assumption that crops would be arranged in rows in the field. [40] train a support vector machine (SVM) based image classifier to recognize four different weed species with a bag of visual words [46] framework. [26] again use a SVM classifier to assign image patches to one of the four classes (background, weed, and two crop classes) based on texture and morphological features. Those methods rely on manual feature extraction, thresholding, and often consist of multi-stage pipelines. We propose an end-to-end trainable approach without any predefined parameters or long pipelines.

More recent approaches deploy deep neural network architectures for different tasks of weed monitoring. [10] focus on real-time mapping of weeds and classify image patches into six classes (crop, soil, and four weed species). However, since the main focus is on the real-time recognition and mapping, requiring fast and memory-efficient computations, the predictions are not fine-grained and one image patch is assigned to the predicted class of the plant in the patch center. [30] do blob-wise classification following a preprocessing step for vegetation detection. Hafiane *et al.* [42] deploy Vision Transformers (ViT) for image classification of weed and crops, outperforming state-of-the-art models when used with small scale datasets. While these methods have good performances for their respective tasks, image classification and object detection fail to address some of the main challenges of automated weed monitoring such as occlusion or overlapping leaves. Moreover, those works test their methods using small-scale self-collected datasets, limiting the comparability of their approaches.

Semantic segmentation enables the differentiation of overlapping organs (i.e. leaves) of crops and weeds and therefore is a more suitable task for effective weed monitoring. In the recent years, various methods utilizing existing semantic segmentation architectures have been proposed for pixel-wise weed segmentation. [20] modify the Mask-RCNN [18] architecture by adding a CBAM (convolutional block attention module) [54] for weed detection in sugarbeet field, and [28] use a fully convolutional SegNet [2] to segment Sagittaria trifolia (a common weed species) in rice fields. Wang *et al.* [49] use Deeplabv3+ [8] and show that segmentation performance can be improved by using near-infrared (NIR) images, especially under difficult lighting conditions. Milioto *et al.* [31] introduce a CNN-based method that allows semantic segmentation of weeds and crops in real-time (20 Hz). They propose an encoder-decoder architecture and use a 14-channel input comprising different vegetation indices and color space values. [58] use both RGB and NIR images and extend a simple fully CNN by adding extra blocks such as UFA (universal function approximation) or attention blocks. Recently, [56] improved semantic segmentation performance using a multi-scale convolutional attention network (MSFCA-Net), and a hybrid dice-focal loss. While these methods have impressive results on semantic segmentation of weeds and crops, they do not report scores on how confident their predictions are, thus the extent those algorithms can be integrated into real-world systems and deployed in real-life scenarios is limited. Our probabilistic method provides uncertainty estimates in addition to segmentation predictions.

2.2. Bayesian Deep Learning

Most deep neural networks (DNNs), which achieve state-of-the-art performance in a variety of computer vision tasks, are deterministic [8, 51, 2, 32]. Typically, they deliver point predictions for quantities of interest, which can lead to overconfident false predictions. This hinders their usage in in many real life applications where reliability, trust and uncertainty around predictions play crucial roles.

Bayesian network architectures assume probability distributions on the model weights, and the aim is to approximate the posterior distribution given the training data. As a result, apart from the final predictions, they can also predict uncertainty estimates. There are multiple approaches to realize Bayesian deep learning in practice, such as Markov chain Monte Carlo (MCMC) methods [52] and deep ensembles [25]. In this work we use variational inference, where the often untractable true posterior is approximated through a tractable variational approximation. The latter is chosen such that the Kullback-Leibler (KL) divergence between the two is minimized [36].

Bayesian DNNs gained interest also in the context of various computer vision tasks, such as segmentation [36, 22] and detection [17], in a variety of application areas like medical image analysis [24, 44] and autonomous driving [29]. Moreover, Bayesian DNNs are shown to perform better compared to their deterministic counterparts [22, 36]. In this work, we also verify this through our experiments.

2.3. Uncertainty Quantification for Semantic Segmentation

Two main uncertainties that are commonly used with Bayesian computer vision models are epistemic and aleatoric uncertainties. Epistemic uncertainty is sometimes also referred to as the model uncertainty. It serves as a measure of how well the model knows about the given input, and hence this uncertainty can be reduced by introducing more training samples. Aleatoric uncertainty on the other hand is referred to as the noise inherent uncertainty. It could be caused by any environmental noise factor (e.g. illumination) or camera noise.

A common practice to quantify confidence (or uncertainty) in deterministic models is to use softmax scores. Not only this would give a false sense of security about the model confidence (as an uncertain output can have a high softmax value [15]), but uncertainty metrics based on softmax cannot capture epistemic uncertainty [37]. In this work, we quantify both uncertainties using entropy based measures.

There is very limited research in probabilistic methods for agricultural phenotyping. In the context of weed detection, Rainville *et al.* [11] use a naive Bayes classifier combined with a Gaussian mixture model to do binary classification on images (crop or weed). However, those probabilistic blocks are at the end of a long feature extraction pipeline that relies on several heuristic parameters. Furthermore, they do not report uncertainty or confidence scores despite being offered by the method in principle. To the best of our knowledge, this paper is the first work proposing uncertainty quantification in the context of weed detection.

3. A Bayesian Segmentian Approach

In this section, we first describe the Bayesian segmentation network used. Then we explain how the posterior predictive distribution is approximated using variational inference. Lastly, we describe the employed uncertainty metrics for semantic segmentation of crops and weeds. The entire model pipeline of our probabilistic approach with uncertainty estimates is illustrated in Figure 2.

3.1. Bayesian DNNs for Semantic Segmentation

Given an input image of size $M \times N$ with F feature dimensions ($F = 3$ if only RGB channels are used, and $F = 4$ if the NIR channel is also added), a semantic segmentation network outputs pixel-level class predictions for C classes ($C = 3$ when only the background (soil), crop, and weed classes are considered). In this work, we deploy

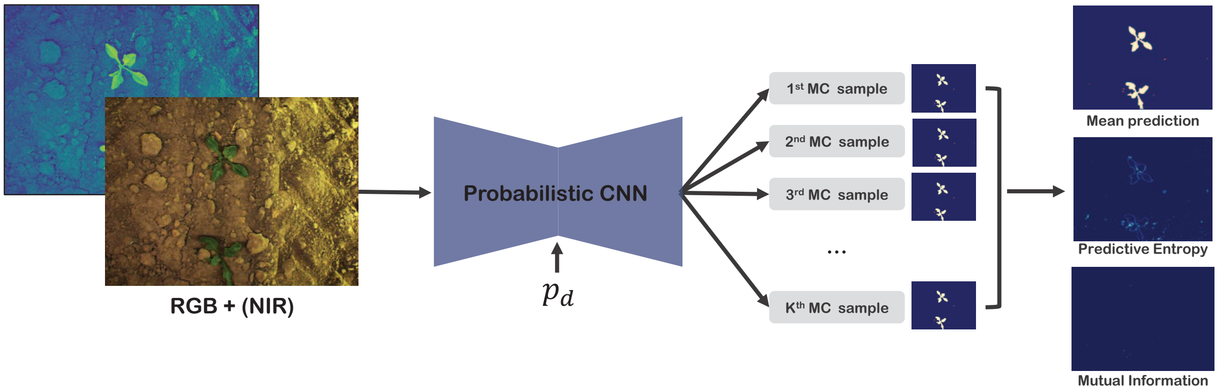


Figure 2: **Overview of the Bayesian segmentation approach**. Given an input RGB image (optionally also with NIR channel), the model predicts segmentation masks for weeds, crops, and soil (background). The segmentation backbone contains dropout layers with a dropout probability of p_d. Those layers are kept on during test time and the model outputs different predictions after K stochastic forward passes. Mean prediction masks and pixel-wise uncertainty estimates are calculated based on these predictions.

DeepLabv3 [7], an established network for semantic segmentation.

In its original form, DeepLabv3 outputs deterministic point estimates. We modify this architecture to obtain its end-to-end trainable probabilistic variant. DeepLabv3 uses dilated convolutions [60, 6] that utilize large receptive fields over the input features. We pair it with ResNet-50 [19] backbone for feature extraction, and extend the architecture into its probabilistic variant by introducing dropout layers. Ideally, a dropout layer should be added after every convolutional layer [22], but this would introduce too much stochasticity and would make convergence very difficult. Therefore we add a dropout layer after every set of residual blocks for the first three sets.

Instead of point estimates, Bayesian deep neural networks model weights as probability distributions. The resulting posterior predictive distribution is given as:

$$p(y^*|x^*, D_n) = \int p(y^*|x^*, w)p(w|D_n)\,dw \tag{1}$$

where x^* and y^* are the input and prediction during test time respectively, w corresponds to set of learnable network parameters, D_n is the training dataset that consists of n pairs (x_i, y_i) of training data with inputs x_i and outputs y_i. We use the posterior predictive means to obtain final point segmentation masks for soil, crops and weeds, and uncertainty measures are derived from variational inference as described next.

3.2. Variational Inference

The computationally limiting part in Equation 1 is the computationally intractable posterior distribution $p(w|D_n)$. This is because the dimension of w is typically large. However, the problem becomes tractable through variational inference. Here, rather than trying to match the true posterior exactly, the latter is approximated by a tractable but close enough approximation q, called the variational approximation. To measure closeness between q and $p(w|D_n)$, the reverse KL divergence is usually used. It can be shown that minimizing the KL divergence $\mathrm{KL}(q(w)||p(w|D_n))$ is equivalent to maximizing the evidence lower bound (ELBO) [38], which leads to a tractable optimization through commonly used loss functions (e.g. cross entropy loss) and stochastic optimizers [14].

In our case, $q(w)$ is a Bernoulli distribution, where nodes in the dropout layer are randomly kept on or off with the given probability p_d. We achieve this by adding dropout layers through MC dropout. Adding these dropout layers also serves as a measure of regularization which helps to stabilize estimates [47]. Importantly, the dropout layers are also kept on during test time. Through multiple forward passes with dropout layers during inference, a predictive probability distribution is obtained.

3.3. Uncertainty Quantification

As proposed by Mukhoti and Gal [36], we use predictive entropy (PE, captures both aleatoric and epistemic uncertainty) and mutual information (MI, captures epistemic uncertainty) to quantify epistemic and aleatoric uncertainties of our probabilistic semantic segmentation model.

Entropy of a discrete random variable X with probability distribution $p(x)$ is defined as:

$$H(x) = -\sum_{x\in\mathcal{X}} p(x)\log(p(x)) \tag{2}$$

From this, PE for the semantic segmentation task with C classes is defined as:

$$H[y^*|x^*, D_n] = -\sum_{c\in\mathcal{C}} \left(\frac{1}{K} \sum_{k=1}^{K} p(y^* = c|\mathrm{x}^*, w_k) \cdot \log \left(\frac{1}{K} \sum_{k=1}^{K} p(y^* = c|\mathrm{x}^*, w_k) \right) \right), \tag{3}$$

where $\mathcal{C} = \{0, 1, \ldots, C-1\}$ with C possible classes, K is the number of stochastic forward passes repeated during test time, and x^* and y^* are the given input sample and prediction variable during test time respectively. MI between the posterior and predictive distribution is given as [36]:

$$I[y^*, w|x^*, D_n] = H[y^*|x^*, D_n] + \frac{1}{K} \sum_{k} \sum_{c} p(y^* = c|x^*, w_k) \log p(y^* = c|x^*, w_k) \tag{4}$$

For an input image of size $M \times N$, the aggregated tensor containing softmax prediction probabilities for C classes after K stochastic passes, is of size $K \times C \times M \times N$. Therefore, similar to obtaining segmentation masks (i.e. each pixel is assigned to the class with the highest softmax score), PE and MI are also calculated for every pixel. As a result, $M \times N$ uncertainty maps are obtained.

4. Experiments

4.1. Dataset and Evaluation Metrics

Dataset. We test our method on the publicly available Sugarbeets2016 [5] dataset. Sugarbeets2016 contains RGB and NIR images that are obtained with a field robot in a sugar beet field in Bonn, Germany, along with pixel-level segmentation masks for three classes: background, crop, and weed. The data is acquired through the growing season of spring 2016, therefore contains images from different growth stages. Images have the size of 1296 × 966 pixels, corresponding to millimeter level resolution. Because the dataset doesn't have an official training-validation-test-split, we choose a random subset of 3,858 images from the dataset for our experiments, and do a random split of 75%-15%-15% for training, validation, and test sets respectively.

Metrics. To evaluate segmentation performance, we use the IoU (intersection-over-union between ground truth and predicted pixels) metric, which is commonly used for semantic segmentation tasks [9, 63, 8]. We report class IoU scores for all three classes, as well as the mIoU (IoU averaged over all classes). Secondly, we evaluate uncertainty of predictions to quantify in which cases the model is confident and in which it is not. In addition, we are interested in evaluating whether uncertainty measures around point predictions are calibrated and sharp and thus hold the potential to be leveraged to make better informed decisions, rather than only relying on the segmentation point prediction. To do so, we report accuracy-uncertainty (AU) maps proposed by Mukhoti *et al.* [36]. In this approach, predicted masks and uncertainties (predictive entropy scores) are considered together and grouped into the following four categories: (1) Accurate and certain (ac), (2) Accurate and uncertain (au), (3) Inaccurate and certain (ic) and (4) Inaccurate and uncertain (ic). Ideally, a well calibrated model would score high in ac and iu. Pixels that are correctly predicted are considered accurate, and pixels that have low uncertainty are considered certain. Computing those metrics individually for every pixel ignores the regional information, therefore they are calculated considering all pixels within a window of window size w_q. The thresholds for deciding whether a pixel is accurate or certain are t_a and t_c and these are applied to the average score of the window. Afterwards the map is upsampled to the original image size, where every pixel inside the patch is assigned the same value.

4.2. Implementation Details

We use ResNet50 [19] as the backbone for the DeepLabv3 architecture, and modify the first convolutional layer to fit to our 4-channel input. The dropout probability is set to $p_d = 0.5$. The Bayesian DeepLabv3 is trained using the ADAM optimizer [23] with a learning rate of 0.001, $\beta_1 = 0.9$ and $\beta_2 = 0.999$. Standard data augmentation (i.e. rotation, flipping, random jitter) is applied to reduce overfitting. Due to the class imbalance between soil and vegetation pixels, we used a weighted cross-entropy loss with the weights 0.2, 0.8, 0.8 for background, crop and weed respectively. During test time, we do 5 stochastic forward passes (i.e. $K = 5$). We train the model for 160 epochs with batch size 4 with an Nvidia A40 GPU and an Intel Xeon Ice Lake CPU with 32 GB memory, which takes approximately 44 hours. Settings have been chosen based on the validation dataset.

4.3. Comparison with the State of the Art

This section compares the segmentation performance of our method denoted by Bayesian DeepLabv3 to three state-of-the-art methods for semantic segmentation using the Sugarbeets2016 dataset. These are CNN-UFAB [59], RSS [31] and MSFCA-Net [57], and the results are taken from [59], [31], and [57] respectively. Since Sugarbeets2016 doesn't have an official test set, each method uses random splits. MSFCA-Net [57] uses 2,677 randomly selected images, divided into 70%-20%-10% train-validation-test split. [31] uses 10,036 images with a 70%-15%-15% split. Lastly, [59] uses 9,070 images with a 80%-20% train-test split.

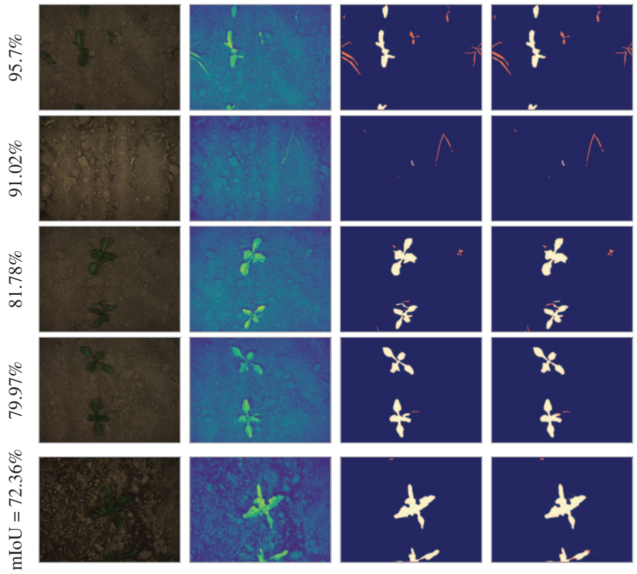
The scores, as well as the input modalities used by each method can be seen in Table 1. Our Bayesian DeepLabv3 approach achieves the highest IoU scores for background

Method	Input	Mode	IoU_{bg}	IoU_{weed}	IoU_{crop}	mIoU
CNN-UFAB [59]	RGB+NIR	D	99.73	**75.26**	92.04	89.01
RSS [31]	14 channels	D	99.48	59.17	83.72	80.80
MSFCA-Net [57]	RGB	D	99.79	73.32	95.62	**89.58**
Bayesian DeepLabv3	RGB+NIR	P	**99.93**	69.31	**95.89**	88.37

Table 1: **Segmentation scores (%) on the Sugarbeets2016 [5] test set.** Input represents the modality of the input images (except RSS [31] since it uses 14 pre-processed channels). The network mode is either deterministic (D) or probabilistic (P). Scores of existings methods are taken from [59, 31, 57].

and crop classes. For weed segmentation, even though not the highest performing one, our method still achieves a competitive score with 5.95% difference with the best method CNN-UFAB. Background is almost perfectly predicted by all methods, though ours scores 0.2% higher. Moreover, similar to [59], we notice many misannotated samples in the Sugarbeets2016 dataset, which naturally leads to lower scores than the actual performance.

4.4. Qualitative Results



(a) Input (RGB) **(b)** Input (NIR) **(c)** Ground truth (d) Prediction

Figure 3: Qualitative results of semantic segmentation of crops (yellow), weeds (orange) and soil (blue) on the Sugarbeets2016 [5] test set. mIoU scores are on the left.

Qualitative results on are exemplified in Figure 3, which depicts two high and three poor scoring examples. It can be seen that the main bodies of both types of vegetation (crops and weeds) are almost correctly segmented. The performance is lower at the tips and thinner crop leaves. This is expected, since these are where crops and weeds look the most alike. Failure cases often happen in the occurrence of occlusion or when leaves of crops and weeds touch each other (which naturally appears as a connected object in a 2D image). Images in this dataset are collected during rather early growth stages of the plants, therefore leaf coverage is low and plants are small. Weeds are significantly smaller than crops. It can furthermore be observed that due to this imbalance, even when a small area of weed is misclassified, the overall prediction score significantly drops.

4.5. Uncertainty Quantification

Figure 4 shows predictive entropy and mutual information, along with AU maps (each pixel is assigned to one of the four cases: ac, au, ic, iu) for images from Sugarbeets2016 test set. Both, t_c and t_a are set to 0.5 (i.e. if more than half of the pixels in the window are correctly predicted, that patch is considered accurate), and w_q is set to 3. AU maps are computed based on predictive entropy.

We make the following observations: Uncertainty is high at the edges of plants. This is clearly seen from the entropy map and the strong presence of both au and ic around the contours. This is because of the fact that PE is a measure of aleatoric uncertainty, and noise is expected at the object edges [36, 41]. On the other hand, mutual information is more concentrated inside the objects (i.e. inside the leaves). This is because MI captures epistemic uncertainty, and high epistemic uncertainty implies the model doesn't know much about that object. This applies to leaves of sugar beet which are thinner and less curvy, therefore the model is unsure about them. The relationship between prediction performance and uncertainty is also observed: The prediction score is higher for the AU maps with fewer au and ic pixels. iu pixels hold potential to be used to improve prediction scores, because those are the pixels the model predicts wrong, but knows that it is uncertain about the prediction.

4.6. Sensitivity Analysis and Comparison with Deterministic Models

To perform sensitivity analyses and validate our choice of using RGB+NIR input configuration, we use both DeepLabv3 and UNet [43] architectures. To make UNet probabilistic, we insert dropout layers after each encoder and decoder unit in the deepest half of the network, as this is found to be the optimal configuration by [22]. We train both architectures with the aforementioned dataset and training configurations, except data augmentation to exemplify that data augmentation can further improve the scores. We use three different input configurations (i.e. RGB, NIR, or 4-channel RGB+NIR). Furthermore, we repeat the same experiments for the deterministic variants of the respective network architectures. The scores are reported in Table 2. RGB+NIR leads to best segmentation performance, while

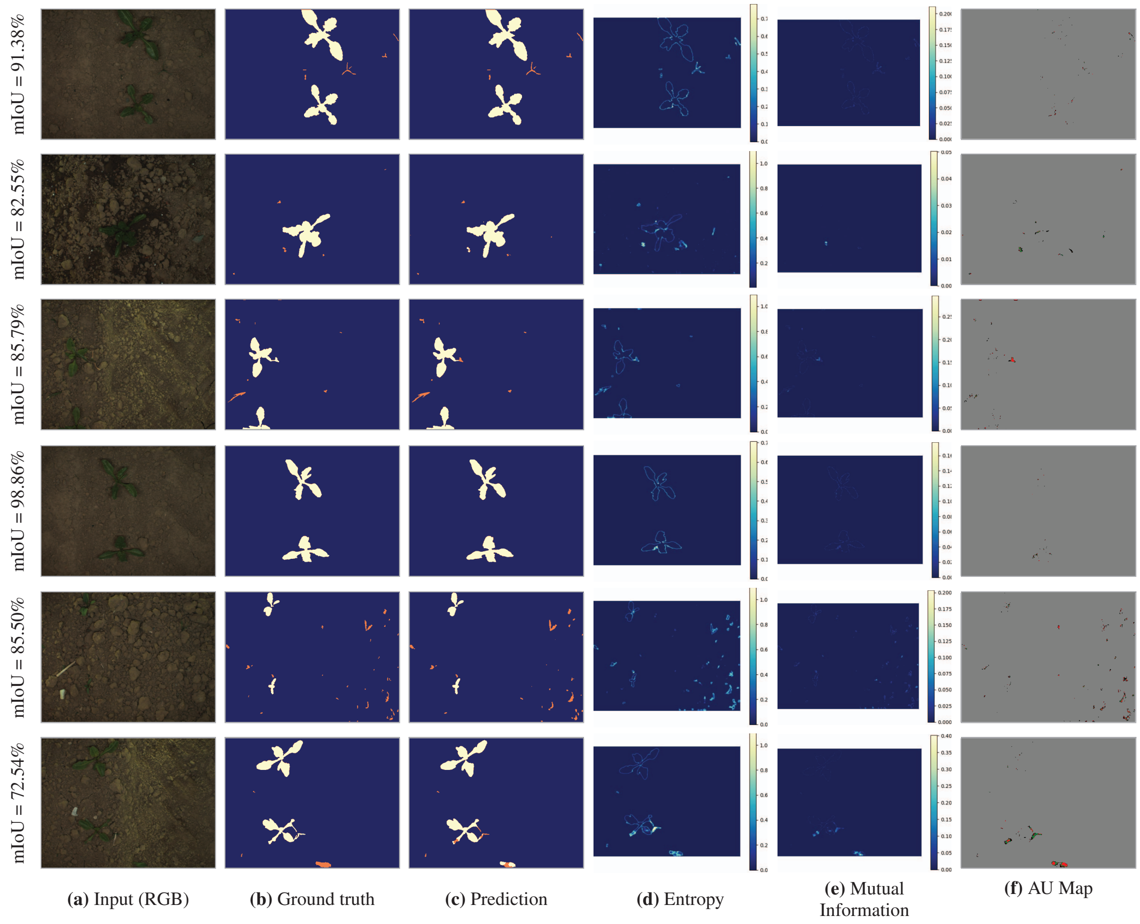


Figure 4: **Uncertainty quantification and semantic segmentation on Sugarbeets2016 [5] test set.** Semantic segmentation classes are crops (yellow), weeds (orange) and soil (blue). The accuracy-uncertainty (AU) map shows ac (gray), au (black), ic (red), and iu (green) pixels together. Best viewed on screen zoomed in due to thin plant structures and edge uncertainties.

	Input	Mode	IoU_{weed}	IoU_{crop}	mIoU
UNet	RGB	D	41.44	92.90	78.08
		P	43.02	92.77	78.57
	RGB+NIR	D	48.07	93.78	80.59
		P	50.35	93.33	81.19
	NIR	D	31.56	90.35	73.91
		P	32.84	90.47	74.38
DLv3	RGB	D	46.36	90.79	79.00
		P	45.98	90.98	78.94
	RGB+NIR	D	62.12	94.95	85.66
		P	62.73	95.00	85.89
	NIR	D	37.90	93.12	76.96
		P	38.60	93.28	77.25

Table 2: Comparison of deterministic (D) and probabilistic (P) variants of UNet [43] and DeepLabv3 [7] with different input modalities on Sugarbeets2016 [5] validation set.

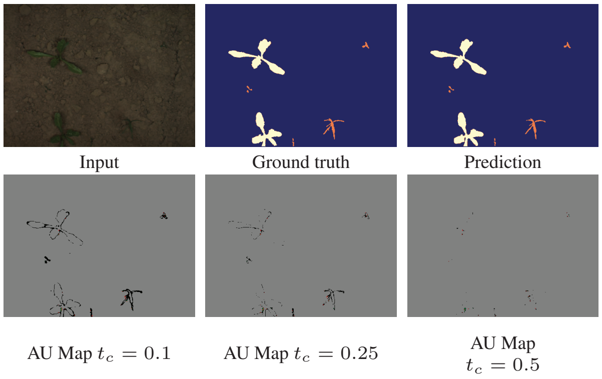
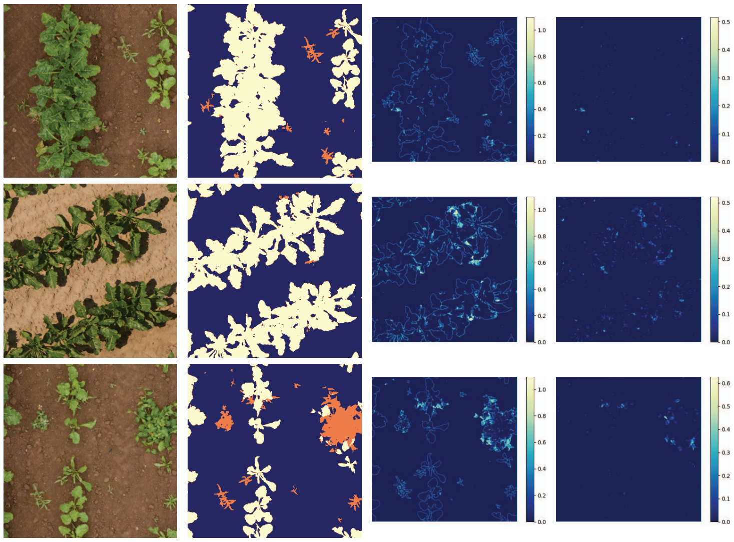


Figure 5: Accuracy-uncertainty maps (ac (gray), au (black), ic (red), and iu (green)) for different certainty thresholds t_c. Best viewed on screen, zoomed in.

using only NIR results in significantly lower scores. The probabilistic variants perform better than their deterministic



(a) Input **(b)** Prediction **(c)** Entropy **(d)** Mutual Information

Figure 6: Semantic segmentation of crops (yellow), weeds (orange) and soil (blue), and uncertainty predictions on PhenoBench [53] test set.

Method	Input	IoU_{bg}	IoU_{weed}	IoU_{crop}	mIoU
Bayesian DeepLabv3	RGB	99.33	63.37	94.60	85.77

Table 3: Results on PhenoBench [53] validation set.

counterparts, except the DeepLabv3-RGB configuration.

As for uncertainty quantification, AU maps for different certainty thresholds are shown in Figure 5. As it would be expected, as the threshold for ”being certain” is lowered, the AU maps become fuller. The difference is most visible with au, meaning that accurately segmented pixels tend to move to the uncertain group with lower accuracy thresholds.

5. Application on PhenoBench Dataset

In this section, we apply our method to the PhenoBench [53] dataset, which is a novel dataset for semantic interpretation of crops and weeds. Compared to the ground vehicle setup of Sugarbeets2016 [5], images for PhenoBench are acquired with a UAV. It contains 2,872 RGB images of sugar beet plants and weeds, with pixel-level annotations at different levels of detail such as semantic or leaf instance masks. PhenoBench has an official training-validation-test split, but labels for the test set are not publicly available. Hence, we report scores on the validation set. We train our model on the PhenoBench training set with the same configuration as in Section 4.2. Prediction scores on validation set and qualitative predictions with uncertainty estimates on the test set can be seen from Table 3 and Figure 6 respectively.

6. Conclusion

We have presented, for the first time, an approach to use Bayesian DNNs for semantic segmentation and uncertainty quantification for weeds and crops. Specifically, we have implemented the probabilistic variant of the DeepLabv3 [7] architecture, and tested it on Sugarbeets2016 dataset [5]. The results demonstrate that the Bayesian model achieves competetive segmentation performance, and in addition, outputs uncertainty maps based on predictive entropy and mutual information, that highlight areas where the model is most uncertain about. We show that uncertainty quantification can be used to shed light on the reliability of a prediction, which makes our approach suitable to be integrated into real-world automated weeding systems. Moreover, we show the applicability of our method to different data acquisition techniques (i.e. field robot and drone) by applying it to the novel PhenoBench [53] dataset. In future work, we plan to carry out extensive experiments on PhenoBench to compare our approach to the recently released benchmark in terms of semantic segmentation performance. Furhermore, we hope to extend our approach to different architectures and multi-temporal datasets, and eventually explore other remote sensing data such as hyperspectral or 3D images.

7. Acknowledgements

The authors acknowledge the support of the Helmholtz Einstein International Berlin Research School in Data Science (HEIBRiDS).

References

[1] Alicia Allmendinger, Michael Spaeth, Marcus Saile, Gerassimos G. Peteinatos, and Roland Gerhards. Precision chemical weed management strategies: A review and a design of a new cnn-based modular spot sprayer. *Agronomy*, 12(7), 2022.

[2] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2481–2495, 2017.

[3] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, Oct 2001.

[4] Chung-Liang Chang and Kuan-Ming Lin. Smart agricultural machine with a computer vision-based weeding and variable-rate irrigation scheme. *Robotics*, 7(3):38, 2018.

[5] Nived Chebrolu, Philipp Lottes, Alexander Schaefer, Wera Winterhalter, Wolfram Burgard, and Cyrill Stachniss. Agricultural robot dataset for plant classification, localization and mapping on sugar beet fields. *The International Journal of Robotics Research*, 36(10):1045–1052, 2017.

[6] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):834–848, 2018.

[7] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.

[8] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.

[9] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3213–3223, 2016.

[10] Tibor de Camargo, Michael Schirrmann, Niels Landwehr, Karl-Heinz Dammer, and Michael Pflanz. Optimized deep learning model as a basis for fast uav mapping of weed species in winter wheat crops. *Remote Sensing*, 13(9):1704, 2021.

[11] François-Michel De Rainville, Audrey Durand, Félix-Antoine Fortin, Kevin Tanguy, Xavier Maldague, Bernard Panneton, and Marie-Josée Simard. Bayesian classification and unsupervised learning for isolating weeds in row crops. *Pattern Analysis and Applications*, 17:401–414, 2014.

[12] Huimin Fang, Mengmeng Niu, Xinzhong Wang, and Qingyi Zhang. Effects of reduced chemical application by mechanical-chemical synergistic weeding on maize growth and yield in east china. *Frontiers in Plant Science*, 13, 2022.

[13] Food and Agriculture Organization. How to feed the world in 2050, 2020.

[14] Yarin Gal and Zoubin Ghahramani. Bayesian convolutional neural networks with bernoulli approximate variational inference. *arXiv preprint arXiv:1506.02158*, 2016.

[15] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1050–1059, 2016.

[16] Roland Gerhards, Dionisio Andújar Sanchez, Pavel Hamouz, Gerassimos G. Peteinatos, Svend Christensen, and Cesar Fernandez-Quintanilla. Advances in site-specific weed management in agriculture—a review. *Weed Research*, 62(2):123–133, 2022.

[17] Ali Harakeh, Michael Smart, and Steven L Waslander. Bayesod: A bayesian approach for uncertainty estimation in deep object detectors. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 87–93, 2020.

[18] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017.

[19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[20] Shangzhu Jin, Haojun Dai, Jun Peng, Yuanmin He, Min Zhu, Wencheng Yu, and Qingxia Li. An improved mask r-cnn method for weed segmentation. In *IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pages 1430–1435, 2022.

[21] R. E. Jones, D. T. Vere, Y. Alemseged, and R. W. Medd. Estimating the economic cost of weeds in australian annual winter crops. *Agricultural Economics*, 32(3):253–265, 2005.

[22] Alex Kendall, Vijay Badrinarayanan, and Roberto Cipolla. Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding. *arXiv preprint arXiv:1511.02680*, 2016.

[23] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015.

[24] Yongchan Kwon, Joong-Ho Won, Beom Joon Kim, and Myunghee Cho Paik. Uncertainty quantification using bayesian neural networks in classification: Application to biomedical image segmentation. *Computational Statistics & Data Analysis*, 142:106816, 2020.

[25] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in Neural Information Processing Systems*, 2017.

[26] Vi Nguyen Thanh Le, Selam Ahderom, and Kamal Alameh. Performances of the lbp based algorithm over cnn models for detecting crops and weeds with similar morphologies. *Sensors*, 20(8), 2020.

[27] Philipp Lottes, Markus Hörferlin, Slawomir Sander, and Cyrill Stachniss. Effective vision-based classification for separating sugar beets and weeds for precision farming. *Journal of Field Robotics*, 34(6):1160–1178, 2017.

[28] Xu Ma, Xiangwu Deng, Long Qi, Yu Jiang, Hongwei Li, Yuwei Wang, and Xupo Xing. Fully convolutional network for rice seedling and weed image segmentation at the seedling stage in paddy fields. *PLOS ONE*, 14(4):1–13, 04 2019.

[29] Rhiannon Michelmore, Matthew Wicker, Luca Laurenti, Luca Cardelli, Yarin Gal, and Marta Kwiatkowska. Uncertainty quantification with statistical guarantees in end-to-end autonomous driving control. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 7344–7350, 2020.

[30] Andres Milioto, Philipp Lottes, and Cyrill Stachniss. Real-time blob-wise sugar beets vs weeds classification for monitoring fields using convolutional neural networks. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4:41–48, 2017.

[31] Andres Milioto, Philipp Lottes, and C. Stachniss. Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in cnns. *IEEE International Conference on Robotics and Automation (ICRA)*, pages 2229–2235, 2017.

[32] Ishan Misra, Rohit Girdhar, and Armand Joulin. An end-to-end transformer model for 3d object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2906–2917, October 2021.

[33] Nur Adibah Mohidem, Nik Norasma Che'Ya, Abdul Shukor Juraimi, Wan Fazilah Fazlil Ilahi, Muhammad Huzaifah Mohd Roslim, Nursyazyla Sulaiman, Mohammadmehdi Saberioon, and Nisfariza Mohd Noor. How can unmanned aerial vehicles be used for detecting weeds in agricultural fields? *Agriculture*, 11(10):1004, 2021.

[34] Nur Adibah Mohidem, Nik Norasma Che'Ya, Abdul Shukor Juraimi, Wan Fazilah Fazlil Ilahi, Muhammad Huzaifah Mohd Roslim, Nursyazyla Sulaiman, Mohammadmehdi Saberioon, and Nisfariza Mohd Noor. How can unmanned aerial vehicles be used for detecting weeds in agricultural fields? *Agriculture*, 11(10):1004, Oct 2021.

[35] António Monteiro and Sérgio Santos. Sustainable approach to weed management: The role of precision weed management. *Agronomy*, 12(1), 2022.

[36] Jishnu Mukhoti and Yarin Gal. Evaluating bayesian deep learning methods for semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018.

[37] Jishnu Mukhoti, Joost van Amersfoort, Philip HS Torr, and Yarin Gal. Deep deterministic uncertainty for semantic segmentation. *arXiv preprint arXiv:2111.00079*, 2021.

[38] John T Ormerod and Matt P Wand. Explaining variational approximations. *The American Statistician*, 64(2):140–153, 2010.

[39] María Pérez-Ortiz, Pedro Antonio Gutiérrez, Jose Manuel Peña, Jorge Torres-Sánchez, César Hervás-Martínez, and Francisca López-Granados. An experimental comparison for the identification of weeds in sunflower crops via unmanned aerial vehicles and object-based analysis. In *Advances in Computational Intelligence*, pages 252–262, 2015.

[40] Michael Pflanz, Henning Nordmeyer, and Michael Schirrmann. Weed mapping with uas imagery and a bag of visual words based image classifier. *Remote Sensing*, 10(10), 2018.

[41] Janis Postels, Francesco Ferroni, Huseyin Coskun, Nassir Navab, and Federico Tombari. Sampling-free epistemic uncertainty estimation using approximated variance propagation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.

[42] Reenul Reedha, Eric Dericquebourg, Raphael Canals, and Adel Hafiane. Transformer neural network for weed and crop classification of high resolution uav images. *Remote Sensing*, 14(3), 2022.

[43] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *arXiv preprint arXiv:1505.04597*, 2015.

[44] Abhinav Sagar. Uncertainty quantification using variational inference for biomedical image segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) Workshops*, pages 44–51, 2022.

[45] Gaurav Sethia, Harish Kumar S. Guragol, Swati Sandhya, J. Shruthi, and N. Rashmi. Automated computer vision based weed removal bot. In *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, pages 1–6, 2020.

[46] Sivic and Zisserman. Video google: A text retrieval approach to object matching in videos. In *Proceedings IEEE international conference on computer vision*, pages 1470–1477. IEEE, 2003.

[47] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958, 2014.

[48] United Nations Department of Economic and Social Affairs. Population, food security, nutrition and sustainable development, 2021.

[49] Aichen Wang, Yifei Xu, Xinhua Wei, and Bingbo Cui. Semantic segmentation of crop and weed using an encoder-decoder network and image enhancement method under uncontrolled outdoor illumination. *IEEE Access*, 8:81724–81734, 2020.

[50] Aichen Wang, Wen Zhang, and Xinhua Wei. A review on weed detection using ground-based machine vision and image processing techniques. *Computers and Electronics in Agriculture*, 158:226–240, 2019.

[51] Wenhai Wang, Jifeng Dai, Zhe Chen, Zhenhang Huang, Zhiqi Li, Xizhou Zhu, Xiaowei Hu, Tong Lu, Lewei Lu, Hongsheng Li, Xiaogang Wang, and Yu Qiao. Internimage: Exploring large-scale vision foundation models with deformable convolutions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14408–14419, 2023.

[52] Max Welling and Yee W Teh. Bayesian learning via stochastic gradient langevin dynamics. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 681–688, 2011.

[53] Jan Weyler, Federico Magistri, Elias Marks, Yue Linn Chong, Matteo Sodano, Gianmarco Roggiolani, Nived Chebrolu, Cyrill Stachniss, and Jens Behley. PhenoBench —

A Large Dataset and Benchmarks for Semantic Image Interpretation in the Agricultural Domain. *arXiv preprint arXiv:2306.04557*, 2023.

[54] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018.

[55] Xiaolong Wu, Stéphanie Aravecchia, Philipp Lottes, Cyrill Stachniss, and Cédric Pradalier. Robotic weed control using automated weed and crop classification. *Journal of Field Robotics*, 37(2):322–340, 2020.

[56] Qiangli Yang, Yong Ye, Lichuan Gu, and Yuting Wu. Msfcanet: A multi-scale feature convolutional attention network for segmenting crops and weeds in the field. *Agriculture*, 13(6):1176, 2023.

[57] Qiangli Yang, Yong Ye, Lichuan Gu, and Yuting Wu. MSFCA-net: A multi-scale feature convolutional attention network for segmenting crops and weeds in the field. *Agriculture*, 13(6):1176, May 2023.

[58] Jie You, Wei Liu, and Joonwhoan Lee. A dnn-based semantic segmentation for detecting weed and crop. *Computers and Electronics in Agriculture*, 178:105750, 2020.

[59] Jie You, Wei Liu, and Joonwhoan Lee. A dnn-based semantic segmentation for detecting weed and crop. *Computers and Electronics in Agriculture*, 178:105750, 2020.

[60] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. In *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, 2016.

[61] Jiacheng Yuan, Jungseok Hong, Junaed Sattar, and Volkan Isler. Row-slam: Under-canopy cornfield semantic slam. In *International Conference on Robotics and Automation (ICRA)*, pages 2244–2250, 2022.

[62] Yangkai Zhang, Mengke Wang, Danlei Zhao, Chunye Liu, and Zhengguang Liu. Early weed identification based on deep learning: A review. *Smart Agricultural Technology*, 3:100123, 2023.

[63] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Scene parsing through ade20k dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 633–641, 2017.