

Real-time detection of grassland weeds

Lukas Petrich^{1*}, Ingo-Leonard Haußmann², Georg Lohrmann²,
Albert Stoll², Volker Schmidt¹

^{1*}Institute of Stochastics, Ulm University, Ulm, D-89069, Germany.

²Hochschule für Wirtschaft und Umwelt Nürtingen-Geislingen,
Neckarsteige 6-10, Nürtingen, D-72622, Germany.

*Corresponding author(s). E-mail(s): lukas.petrich@uni-ulm.de;
Contributing authors: haussmann@hfwu.de; georg.lohrmann@hfwu.de;
albert.stoll@hfwu.de; volker.schmidt@uni-ulm.de;

Abstract

Purpose: This study addresses the challenge of real-time detection of the weeds *Colchicum autumnale* and *Rumex* species on grassland sites, which is an inherently difficult problem because the predominantly green weed leaves provide little contrast to the similarly colored vegetation backgrounds. The resulting detector will be integrated into the SELBEWAG tool, a non-chemical, site-specific weed treatment device.

Methods: We collected and annotated RGB video recordings from grassland sites in Southwest Germany and trained a quantized EfficientDet object detection model, which has been optimized for low latency on edge devices.

Results: The detection system achieved a mean average precision of 0.606 across both weed types (0.617 for *Rumex* and 0.595 for *C. autumnale*). With an optimal decision threshold, the model demonstrated precision values of 56.0% for *C. autumnale* and 48.1% for *Rumex*, with corresponding recall values of 62.1% and 67.1%, respectively. Detection performance was influenced by surrounding vegetation height and weed clustering.

Conclusion: The developed system provides effective real-time detection of grassland weeds suitable for integration with the SELBEWAG tool. While detection challenges remain, in particular in high vegetation conditions, the approach significantly improves upon area-wide treatment methods by targeting approximately twice the necessary area rather than entire fields.

Keywords: weed control, grassland site, machine learning, object detection

1 Introduction

The site-specific control of weeds in grassland sites offers significant advantages over traditional area-wide approaches. This targeted approach can have a positive effect on the productivity of grassland while adhering to nature conservation regulations that apply, e.g., to extensive areas. Despite these benefits, manual weed control remains infeasible due to its high labor intensity, creating a need for automated solutions. The key challenge in automation is the precise localization of weeds, which is essential for effective targeted treatment.

Our research focuses on detection capabilities for the SELBEWAG tool, a site-specific weed control implement. The initial control concept, as described in [Martin et al. \(2022\)](#), features a tractor-mounted device equipped with several high-pressure water jets of 0.25 m width that can be activated independently. This system provides non-chemical treatment of grassland weeds, aligning with environmental sustainability goals and making it particularly suitable for sensitive areas such as water protection zones.

Real-time detection represents a critical advancement over previous approaches. In a previous work, we developed a procedure to detect the weed *Colchicum autumnale* in drone images of grassland sites ([Petrich et al. 2020](#)). While this method offers advantages for optimizing treatment routing, it presents significant disadvantages. The optimal timing for detection and treatment often differs substantially—for example, *Colchicum autumnale* can most easily be detected during its blooming period in autumn, but treatment is most effective in early summer, by which time weed locations may have changed. Additionally, separating detection and treatment into two distinct steps reduces practicality and increases operational effort.

To address these limitations, our approach integrates detection directly on the SELBEWAG treatment tool by leveraging live video streams from rugged cameras mounted at the front of the implement. This integration creates an important trade-off for the weed detection system between accuracy and latency, which directly affects the working speed of the weed control tool. A further constraint is the reliance on less powerful edge devices for running the detection model, necessitating efficient algorithms.

Our research specifically targets two problematic weed types in grasslands: *Colchicum autumnale* and *Rumex* species ([Bown 1995](#)). *C. autumnale*, commonly known as autumn crocus or meadow saffron, presents a serious threat to livestock due to its high toxicity, containing the alkaloid colchicine which can be fatal even in small doses. Its presence in hay or silage can lead to economic losses for farmers and poses animal welfare concerns. *Rumex* species, particularly *R. obtusifolius* or broad-leaved dock, are highly competitive perennials with extensive root systems and prolific seed production, allowing them to rapidly colonize grasslands. They reduce forage quality and yield by displacing valuable grass species, and their low palatability and nutritional value for livestock further diminishes pasture productivity. Traditional site-specific control methods for both weed types are often labor-intensive, highlighting the need for precise, automated detection and treatment systems.

Research on the detection of *C. autumnale* in grasslands is relatively limited, with [Petrich et al. \(2020\)](#) being one of the few studies addressing this specific weed species, as mentioned above. In contrast, the detection of *Rumex* species has been much more

extensively studied in the scientific literature. A comprehensive review by [Binch and Fox \(2017\)](#) provides a thorough comparison of many previous weed detection procedures for Rumex species. The methods examined in this review include among others linear binary patterns, Fourier analysis, and support vector machines, though notably, deep learning approaches were not included, which were less widespread at that time. The considered methods were evaluated on a standardized dataset, providing a fair benchmark. In recent years, deep learning methods have become increasingly prevalent in Rumex detection research. Several studies ([Lam et al. 2021](#); [Valente et al. 2019](#); [Schneider et al. 2022](#); [Valente et al. 2022](#)) have focused on the detection of Rumex using unmanned aerial vehicles (UAV), representing a different approach compared to having the detection system integrated directly into a weed control tool as in our work. [Husham Al-Badri et al. \(2023\)](#) introduced an adaptive non-maximum suppression technique, to eliminate overlapping predicted bounding boxes, and proposed ensemble classifiers with the combination of three extractors at its backbone. For comparative purposes in Section 3, we consider the results of two recent publications: [Güldenring et al. \(2023\)](#) created an open-access Rumex dataset and established a baseline detection model using YOLOX architectures. Similarly, [Heil and Stein \(2024\)](#) compared the effectiveness of RGB and multispectral images for Rumex detection using different YOLO model variants. The present paper differs from previous works in several aspects. First, we address the detection of both *C. autumnale* and Rumex species within a single system. Second, we conduct our evaluation under conditions that closely approximate real-world application scenarios, particularly in the context of the SELBEWAG tool. A preliminary version of our detection model was presented in [Haußmann et al. \(2024a\)](#). The current paper expands upon this earlier work by utilizing an improved dataset and providing more comprehensive details on both the model training and evaluation.

In summary, we present a real-time weed detector capable of identifying *C. autumnale* and Rumex on grassland sites. We aim to integrate this detector into the SELBEWAG weed control tool, and consequently evaluate its capabilities as closely as possible to the real-world application. In particular, this includes design choices and optimizations of the model to achieve low latencies on the target hardware, sometimes at the cost of prediction performance.

2 Materials and methods

2.1 Ground truth datasets

The process of image data acquisition for this study is comprehensively described in [Haußmann et al. \(2024b\)](#). For capturing the grassland imagery, we utilized cameras mounted on an electric wheelbarrow, which effectively simulated the camera configuration of the actual treatment tool in terms of camera model, mounting height, 90° angle to the ground as well as the natural lighting. This setup was chosen for its ease of use and flexibility in field applications. The image data consisted of FullHD (1920 × 1080 pixel) RGB video recordings captured at various grassland sites across South-west Germany. We conducted multiple recordings of the same areas at different times to capture various growth stages of the target weeds and the surrounding vegetation.

For the annotation process, we first divided the source videos into smaller, more manageable video fragments to facilitate efficient labeling. Using the LabelStudio software (Tkachenko et al. 2020), we annotated each target weed with an axis-aligned bounding box. Significant advantages of using video recordings rather than still images were the increased context of having multiple frames of the same area and the ability to interpolate annotations between frames, which reduced the manual labeling effort while maintaining annotation quality.

To prepare the dataset for model training and evaluation, we extracted every third frame from the videos. This sampling rate was determined based on the frame rate of the recordings and the average driving speed of the electric wheelbarrow, ensuring that consecutive extracted frames contained sufficiently different content. We then removed all frames that did not contain any target weed species to focus the dataset on relevant samples. To further refine our dataset, we implemented a perceptual hash-based filtering approach (see below) to eliminate redundant frames. This was particularly important for segments at the beginning of source videos where the camera was activated, but the wheelbarrow remained stationary.

The perceptual image hashing technique (Zauner 2010; Farid 2021) allows identifying different images that appear similar to the human eye. This method is commonly applied in image spam detection and identification of unwanted content such as copyrighted or illegal images. The employed perceptual hashing algorithm operates by first shrinking the input image and converting it to grayscale. It then computes a discrete cosine transform (DCT), which, similar to a Fourier transform, converts the image from a 2D spatial domain to a 2D frequency domain where each value represents the amplitude of a cosine with a given frequency (Burger and Burge 2016). This transformation is also a fundamental component of JPEG compression. The algorithm proceeds by calculating differences between row-wise neighboring “pixels” in the frequency domain. By flattening and concatenating these individual values as binary numbers, we obtain a single binary number—the image hash—that represents the content of original image. To detect similar images, we computed the Hamming distance (i.e. comparing the two binary numbers digit by digit and counting the differences) between the hashes of two given images. If this distance did not exceed a predetermined threshold, the images were considered too similar, and only the first frame of each group of similar frames was retained for each source video.

For the final dataset organization, we assigned each source video to training, validation, and test dataset splits with probabilities of 60%, 20%, and 20%, respectively. Each split then comprised the selected video frames from the corresponding source videos. To optimize the training process, we converted each split into the TFRecord file format, which enables efficient data reading during the model training phase.

Basic statistics of the final ground truth dataset and its splits are shown in Table 1. It can be assumed that the detection performance varies over the course of a season, for example *C. autumnale* often emerges before the surrounding vegetation improving the chances of detection early in the year. For this reason, we also present the dates of the recordings in Figure 1 and provide a detailed evaluation of this phenomenon in Section 3.

dataset split	# videos	# frames	# bounding boxes
training	150	11911	28873
validation	49	5368	14211
test	57	4215	10440
<i>total</i>	256	21494	53524

Table 1: Number of source videos, video frames and bounding boxes in total and for each split of the ground truth dataset.

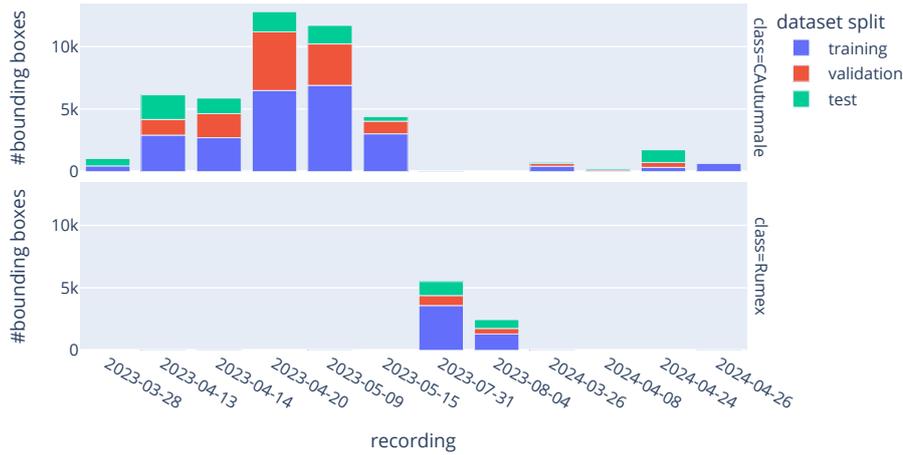


Fig. 1: Number of bounding boxes in the training, validation and test dataset grouped by their recording date. Note that due to their individual availability at the considered grassland sites, each recording focused on a single weed species, but some outliers can still be found (barely visible).

2.2 Detection model

For our weed detection system, we employed the EfficientDet model family (Tan et al. 2020), which represents a one-stage object detector. These architectures utilize the EfficientNetv2 image classification family (Tan and Le 2021) as backbone networks to extract image features. A key component of the EfficientDet architectures is the bidirectional feature network with fast normalization, which enables efficient and effective feature fusion. The model family implements a unified scaling approach, using a single scaling factor to govern the depth, width, and resolution for all backbone, feature, and prediction networks. Based on this scaling methodology, different model architectures are available, ranging from the least to most complex/performant: EfficientDet-D0 through EfficientDet-D7 and EfficientDet-D7x. Our detector implementation is based on the code published by Tan et al. (2020).

For deployment on edge devices such as the computing hardware of the SEL-BEWAG tool, modified models optimized for post-training quantization are available (EfficientDet-lite0 through EfficientDet-lite4). Model quantization (Jacob et al. 2018) involves using efficient 8-bit integer arithmetic instead of 32-bit floating-point numbers during inference, which significantly decreases inference latency, particularly on hardware commonly used for edge devices. While this approach slightly reduces prediction performance, the trade-off is generally favorable for real-time applications. All results presented in this paper were obtained using the quantized model running on the LiteRT (formerly TensorFlow Lite) runtime (Google 2025), ensuring that we report the most realistic prediction performance for practical field applications.

For training our detection model, we utilized pretrained weights from the COCO 2017 dataset (Lin et al. 2014) and employed the Adam optimizer (Kingma and Ba 2015). To enhance model robustness, we implemented image augmentation during training, including horizontal flipping and random resizing of images between 0.8x and 1.2x of the original size before cropping to the original dimensions. These relatively small modifications were appropriate given that the distance from the camera to the ground remains almost constant in our application scenario.

The loss function for our detector combines multiple components as the model outputs a list of bounding boxes and corresponding class scores representing the likelihood that the bounding box belongs to each class (i.e. weed type). For the localization error of the bounding boxes, we employed the Huber loss (Hastie et al. 2009), which is a hybrid between the ℓ_1 and the ℓ_2 error. More specifically, for each bounding box component, the error $\tilde{\ell}_H(\hat{b}, b)$ between the predicted value $\hat{b} \in \mathbb{R}$ and the corresponding true value $b \in \mathbb{R}$ is given by

$$\tilde{\ell}_H(\hat{b}, b) = \begin{cases} \frac{1}{2}(\hat{b} - b)^2, & \text{if } |\hat{b} - b| \leq \delta, \\ \delta|\hat{b} - b| - \frac{1}{2}\delta^2, & \text{otherwise,} \end{cases} \quad (1)$$

for $\delta = 0.1$. For the class error, the focal loss (Lin et al. 2017) with label smoothing was implemented. For each class, consider the predicted score $\hat{c} \in [0, 1]$ for that class and the true value $c \in \{0, 1\}$ being one if and only if the currently considered class is correct. The prediction error $\tilde{\ell}_F(\hat{c}, c)$ is then given by

$$\tilde{\ell}_F(\hat{c}, c) = \begin{cases} \alpha(1 - \hat{c})^\gamma \tilde{\ell}_{CE}(\hat{c}, 1 - \frac{1}{2}\beta), & \text{if } c = 1, \\ (1 - \alpha)\hat{c}^\gamma \tilde{\ell}_{CE}(\hat{c}, \frac{1}{2}\beta), & \text{if } c = 0, \end{cases} \quad (2)$$

where $\alpha = 0.25$, $\gamma = 1.5$, and $\tilde{\ell}_{CE}(\hat{c}, c)$ is the cross entropy loss given by

$$\tilde{\ell}_{CE}(\hat{c}, c) = -c \log \hat{c} - (1 - c) \log(1 - \hat{c}). \quad (3)$$

Here, we incorporated label smoothing with parameter $\beta = 0.1$ to account for potential inaccuracies in the training data. The total loss function aggregates the localization and classification errors over all bounding boxes in a batch of the training data.

The model was trained for 300 epochs with a learning rate of 0.08, except for the initial training step where a learning rate of 0.008 was used. We implemented a cosine

learning rate schedule (Loshchilov and Hutter 2017) to optimize convergence. The loss during training is shown in Figure 2.

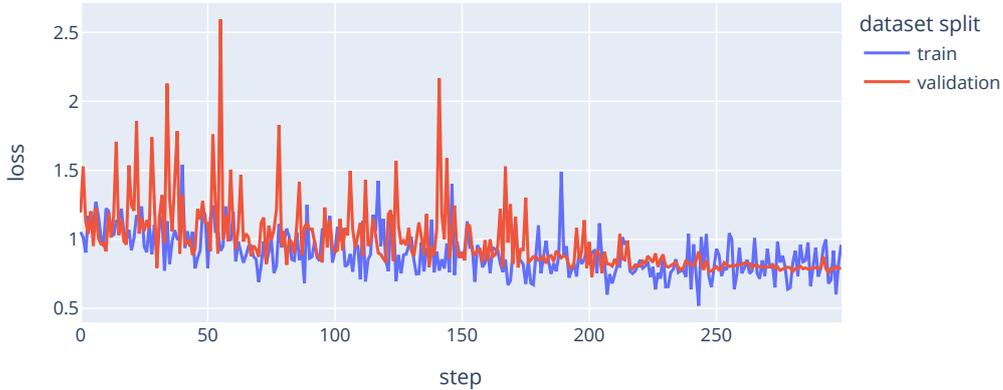


Fig. 2: Total loss during the training of the training and validation dataset.

Through hyperparameter tuning, we found that image size significantly impacts model performance. An input resolution of 768×768 pixels provided the best trade-off between inference latency and prediction performance for our application. Interestingly, the choice of model architecture had less impact on the performance-latency trade-off, leading us to select the smallest architecture, EfficientDet-lite0, for our final implementation.

3 Results and discussion

In our analysis, we consider the output of the detection model as a list of bounding boxes with corresponding class labels (either *C. autumnale* or *Rumex*) and associated confidence scores ranging from 0 to 1. For each detection, only the class label with the highest score was selected. Additionally, we performed non-maximum suppression (NMS, Neubeck and Van Gool (2006)) to prune overlapping bounding boxes, retaining only the one with the highest confidence score. Here, two predicted bounding boxes $B_p^{(1)}, B_p^{(2)} \subset \mathbb{R}^2$ are considered overlapping if $\text{IoU}(B_p^{(1)}, B_p^{(2)}) \geq 0.5$, where $\text{IoU}(B_p^{(1)}, B_p^{(2)})$ is the quotient of the areas of their intersection over union, i.e.,

$$\text{IoU}(B_p^{(1)}, B_p^{(2)}) = \frac{|B_p^{(1)} \cap B_p^{(2)}|}{|B_p^{(1)} \cup B_p^{(2)}|}, \quad (4)$$

where $|B|$ denotes the area of a set $B \subset \mathbb{R}^2$. Given some decision threshold $d_{\text{dec}} \in [0, 1]$, we select only those bounding boxes and corresponding class labels whose confidence scores exceed the threshold d_{dec} .

Similar to well-established evaluation protocols like the one for the COCO competition (Lin et al. 2014), we determine whether a predicted bounding box $B_p \subset \mathbb{R}^2$

correctly identifies a ground truth bounding box $B_{\text{gt}} \subset \mathbb{R}^2$ based on the correct class label and $\text{IoU}(B_{\text{p}}, B_{\text{gt}})$. We employed an IoU threshold of $d_{\text{IoU}} = 0.5$, which is appropriate for our application since localization errors are not critical given that the detected bounding boxes are coarsely mapped to a few nozzles (each with a width of approximately 0.25 m). More precisely, a predicted bounding box B_{p} that overlaps a ground truth bounding box B_{gt} with $\text{IoU}(B_{\text{p}}, B_{\text{gt}}) \geq d_{\text{IoU}}$ is classified as a true positive. If multiple predicted bounding boxes overlap with a single B_{gt} , only the one with the highest IoU is considered a true positive, while the others are classified as false positives (unless they overlap with another ground truth bounding box). Similarly, if a single B_{p} overlaps with multiple ground truth bounding boxes, only the match with the highest IoU is counted as a true positive. A predicted bounding box B_{p} without a matching ground truth bounding box is categorized as a false positive, while a ground truth bounding box B_{gt} without a corresponding predicted bounding box is classified as a false negative.

For our evaluation metrics (Manning et al. 2008), we computed the numbers of true positives l_{tp} , false positives l_{fp} , and false negatives l_{fn} based on the test dataset unless otherwise indicated. The precision $\text{prec} \in [0, 1]$, which represents the probability that a predicted bounding box with a given class correctly predicts a weed of that type, is given by

$$\text{prec} = \frac{l_{\text{tp}}}{l_{\text{tp}} + l_{\text{fp}}}. \quad (5)$$

The recall $\text{rec} \in [0, 1]$, which indicates the probability that a given weed in the ground truth data is correctly predicted by the detector, is defined as

$$\text{rec} = \frac{l_{\text{tp}}}{l_{\text{tp}} + l_{\text{fn}}}. \quad (6)$$

Moreover, we use the $F_{\beta} \in [0, 1]$ metric as an aggregate of precision and recall, given by

$$F_{\beta} = \frac{(\beta^2 + 1) \text{prec} \cdot \text{rec}}{\beta^2 \text{prec} + \text{rec}}, \quad (7)$$

for some $\beta > 0$. Rather than using the more commonly employed F_1 metric, we focus on F_2 , which places greater emphasis on recall than precision. This choice reflects our application priorities: it is more important to detect all weeds (high recall) than to ensure that all predictions are accurate (high precision). In the former case, missing weeds might lead to site repopulation, while in the latter case, only some additional areas might receive unnecessary treatment.

For assessing overall prediction performance, we utilize the precision-recall curve, which plots the maximum precision obtainable for a given recall value. This curve is generated by evaluating the model output at various decision thresholds between 0 and 1. The mean average precision (mAP), which represents the area under the precision-recall curve averaged over all classes, provides a comprehensive scalar performance metric, see Manning et al. (2008) for more details on these metrics.

The precision-recall curve for our model is presented in Figure 3a. The mAP achieved is 0.606, with class-specific values of 0.617 for *Rumex* and 0.595 for *C. autumnale*, indicating similar detection performance for both target weeds. Interestingly,

Rumex detection is slightly better despite having fewer examples in the ground truth dataset (see Section 2.1). This superior performance is likely attributable to the more distinct leaf shapes and coloring of Rumex species. These results for detecting Rumex significantly exceed the $mAP = 0.528$ (IoU threshold of 0.5) obtained in [Güldenring et al. \(2023\)](#) with a model architecture of comparable size (YOLOX-tiny). In the most recent study by [Heil and Stein \(2024\)](#), $mAP = 0.583$ (YOLO-8n) and $mAP = 0.629$ (YOLO-9t) were reported for real-time capable architectures on RGB images. This is very similar to what our model achieved. Note, however, that we additionally performed model quantization to evaluate the most realistic prediction performance, which slightly degrades the results.

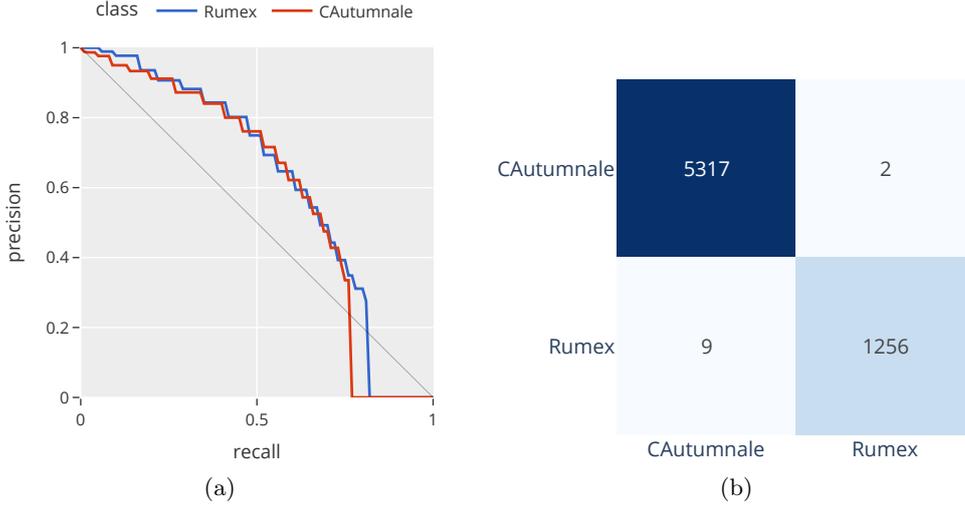


Fig. 3: Precision-recall curve (a) and confusion matrix (b) of the presented weed detector.

The optimal decision threshold d_{dec} —which is necessary for practical application to decide if a detected area should be treated—was computed by maximizing the F_2 metric averaged over all classes. This estimation, performed on the training dataset, yielded a decision threshold of $d_{dec} = 0.213$. Alternative strategies could be employed, for example, to place even greater emphasis on recall.

Example detection results are shown in Figure 4. These images highlight the challenging nature of the detection task, even for human observers, as it involves identifying (mostly) green leaves against a green background. The difficulty is particularly pronounced when surrounding vegetation is high (Figures 4b,e,f) and the target weeds may be partially occluded. Nevertheless, the detector generally performs well, correctly identifying most weeds with only two clear false negatives (Figures 4a,e). Furthermore, some false positives are observed, such as in Figure 4f where the predicted bounding box in the bottom left very likely corresponds to a *Taraxacum* species. In Figure 4c, on the other hand, it is unclear from the image whether there is an annotation error

or indeed another plant such as a *Taraxacum* species. Weed clusters, as exemplified in Figure 4e, present a particular challenge. These clusters often result in many overlapping detections for the same weed, leading to numerous false positives in the evaluation since only the detection with the highest IoU is considered correct. However, these multiple overlapping predictions have minimal impact on practical weed control since the entire area is treated regardless. Moreover, it is often not obvious what the precise extent of each individual weed should be, and, as mentioned above, precise localization is of lower priority given the width of the individually controllable segments of the SELBEWAG tool (approximately 0.25 m).

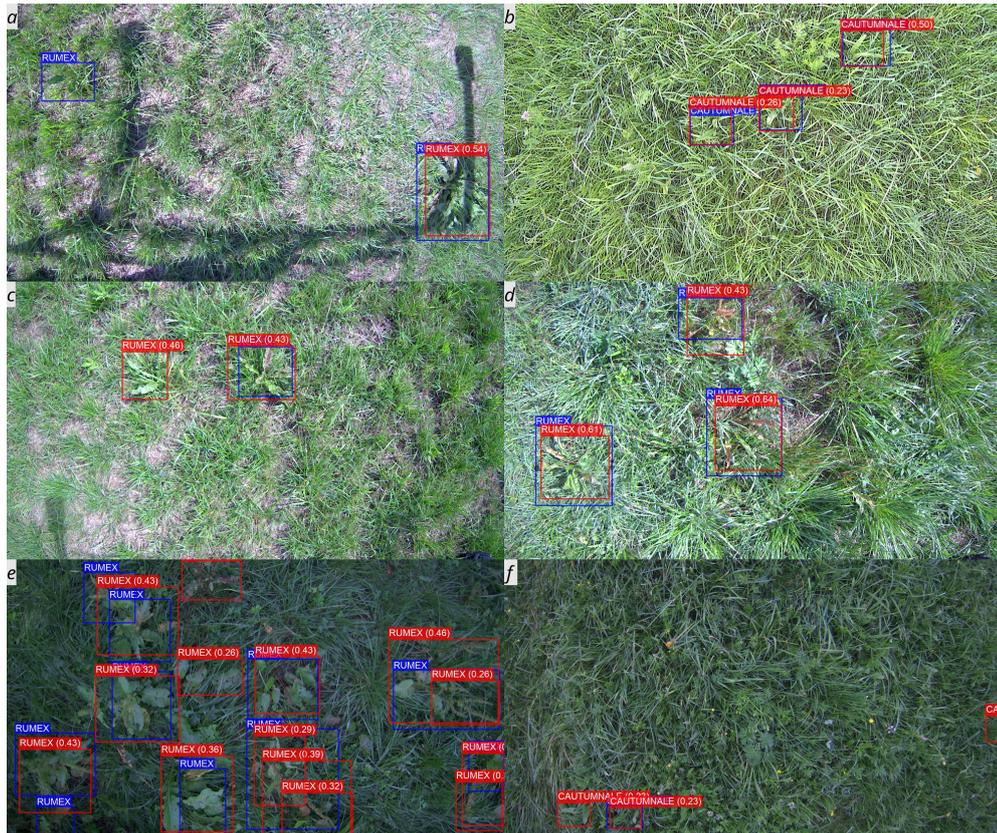


Fig. 4: Randomly selected example images from the test dataset. Ground truth bounding boxes are colored in blue, whereas predicted bounding boxes are indicated in red together with their confidence score.

The confusion matrix presented in Figure 3b reveals very few instances where weed types were misclassified. This is unsurprising given the significant visual differences between *C. autumnale* and *Rumex* species.

The evaluation metrics are summarized in Table 2. The results indicate that 56.0% of *C. autumnale* detections are correct, while the precision for *Rumex* is about 8 percentage points lower. The recall values are better for both weed types, which is by design since we maximized F_2 . As observed in Figure 4, weed detection in grassland represents a challenging problem. High surrounding vegetation often poses challenges for detection accuracy. In general, many false positives for *Rumex* originated from localization problems (see Figure 4e), while others were due to other plants with red-brown leaves. The detector also frequently struggles with different plants that have leaves similar to *C. autumnale*. In practice, however, weeds are usually relatively rare on sites where site-specific weed control is considered compared to the whole area. Our approach results in treating roughly twice the necessary area, which is still significantly better than area-wide weed control.

	prec	rec	F_1	F_2	support
<i>C. autumnale</i>	0.560	0.621	0.589	0.607	8568
<i>Rumex</i>	0.481	0.671	0.560	0.622	1872
<i>micro avg</i>	0.543	0.630	0.583	0.610	10440
<i>macro avg</i>	0.521	0.646	0.575	0.616	10440

Table 2: Evaluation metrics for each weed type as well as their per-instance (micro avg) and per-class averages (macro avg) from the test dataset.

We also investigated how the precision for *C. autumnale* detection varies over time, as detection performance is influenced by surrounding vegetation. Since *C. autumnale* often experiences accelerated growth in late spring/early summer, it might be expected that the timing of recordings affects precision—a consideration important for practical application. Figure 5 illustrates this relationship for *C. autumnale*.

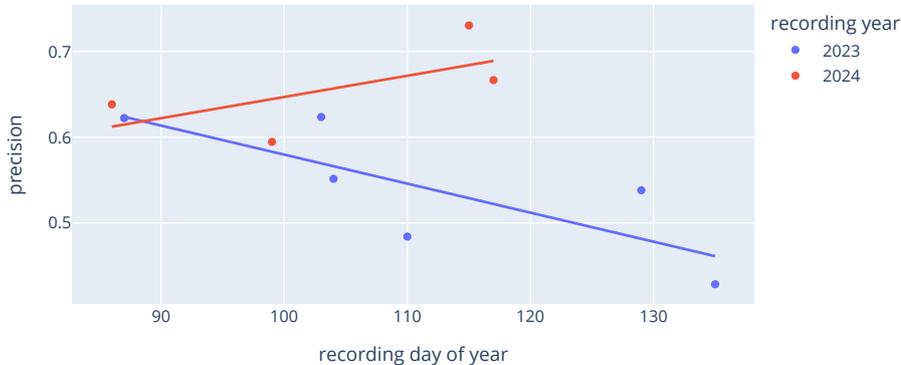


Fig. 5: Precision of the *C. autumnale* detection over time. The precision was estimated for all images of the test dataset that were recorded at the given dates.

For Rumex, too few recording dates were available to conduct a similar investigation. The data for 2023 shows a clear downward trend in precision over time, while the 2024 data exhibits an upward trend, though no late-season recordings (which showed the strongest performance degradation in 2023) were made in 2024. Moreover, recall from Figure 1 that most examples in the ground truth dataset were captured in 2023 providing more weight to the former hypothesis. The differences might also be caused by the different growth conditions such as weather patterns. In summary, the results regarding temporal effects are not definitive, and more data is required to draw conclusions.

4 Conclusion

In this study, we developed and evaluated a real-time detection system for *Colchicum autumnale* and Rumex species in grassland environments, designed specifically for integration with the SELBEWAG non-chemical weed control tool. Our approach successfully addresses the challenging task of identifying predominantly green weeds against similar backgrounds. The quantized EfficientDet model achieved satisfactory detection performance while meeting the operational constraints of edge devices.

Despite the inherent difficulties in weed detection—particularly in areas with high surrounding vegetation—our system significantly improves upon conventional area-wide treatments by targeting approximately twice the necessary area rather than entire fields. The prioritization of recall over precision in our model design ensures that the majority of weeds are detected, preventing site repopulation while still substantially reducing the treated area compared to traditional methods. This selective approach in combination with the non-chemical treatment of the SELBEWAG tool aligns with environmental sustainability goals and is particularly valuable in sensitive areas such as water protection zones.

Future work should focus on further improving detection accuracy in challenging conditions and investigating the temporal effects on detection performance throughout the growing season. Moreover, the presented approach should be extended to further (grassland) weeds. Nevertheless, the current system represents a practical and effective solution for environmentally sensitive grassland management that balances ecological considerations with agricultural productivity. Regarding the integration into the SELBEWAG tool, in an upcoming study we analyze the timing and latencies of the detection and the overall control system.

Declarations

Funding

The project is supported by funds of the Federal Ministry of Food and Agriculture (BMEL) based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the innovation support program.

Competing Interests

The authors have no competing interests to declare that are relevant to the content of this article.

References

- Binch, A. and Fox, C. (2017). Controlled comparison of machine vision algorithms for Rumex and Urtica detection in grassland. *Computers and Electronics in Agriculture*, 140:123–138.
- Bown, D. (1995). *Encyclopedia of Herbs and Their Uses*. DK Publishing.
- Burger, W. and Burge, M. J. (2016). *Digital Image Processing*. Springer.
- Farid, H. (2021). An overview of perceptual hashing. *Journal of Online Trust and Safety*, 1:1–22.
- Google (2025). LiteRT. <https://ai.google.dev/edge/litert>. last accessed 2025-03-25.
- Güldenring, R., van Evert, F., and Nalpantidis, L. (2023). RumexWeeds: A grassland dataset for agricultural robotics. *Journal of Field Robotics*, 40:1639–1656.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. Springer, 2 edition.
- Haußmann, I.-L., Lohrmann, G., Stoll, A., Petrich, L., and Schmidt, V. (2024a). Implementation of a real-time plant detector for a selective grassland weeding machine using high-pressure water jets. In VDI Wissensforum GmbH, editor, *LAND.TECHNIK 2024*, pages 297–302.
- Haußmann, I.-L., Petrich, L., Lohrmann, G., Schmidt, V., and Stoll, A. (2024b). Bilddatenakquisition zur Entwicklung eines Machine-Learning-Detektors für Grünlandunkräuter [Image acquisition for the development of a machine learning detector for grassland weeds]. In Hoffmann, C., Stein, A., Gallmann, E., Dörr, J., Krupitzer, C., and Floto, H., editors, 44. *GIL - Jahrestagung, Biodiversität fördern durch digitale Landwirtschaft*, pages 287–292. Gesellschaft für Informatik e.V.
- Heil, J. and Stein, A. (2024). Detection of Rumex spp. in grassland using multispectral images and deep learning. In VDI Wissensforum GmbH, editor, *LAND.TECHNIK 2024*, pages 223–228.
- Husham Al-Badri, A., Azman Ismail, N., Al-Dulaimi, K., Ahmed Salman, G., and Sah Hj Salam, M. (2023). Adaptive non-maximum suppression for improving performance of Rumex detection. *Expert Systems with Applications*, 219:119634.

- Jacob, B., Kligys, S., Chen, B., Zhu, M., Tang, M., Howard, A., Adam, H., and Kalenichenko, D. (2018). Quantization and training of neural networks for efficient integer-arithmetic-only inference. In O’Conner, L., editor, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2704–2713. IEEE.
- Kingma, D. P. and Ba, J. (2015). Adam: A method for stochastic optimization. In Bengio, Y. and LeCun, Y., editors, *3rd International Conference on Learning Representations, ICLR 2015*.
- Lam, O. H. Y., Dogotari, M., Prüm, M., Vithlani, H. N., Roers, C., Melville, B., Zimmer, F., and Becker, R. (2021). An open source workflow for weed mapping in native grassland using unmanned aerial vehicle: using Rumex obtusifolius as a case study. *European Journal of Remote Sensing*, 54:71–88.
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). Focal loss for dense object detection. In O’Conner, L., editor, *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2999–3007. IEEE.
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. In Fleet, D., Pajdla, T., Schiele, B., and Tuytelaars, T., editors, *Computer Vision – ECCV 2014*, pages 740–755. Springer.
- Loshchilov, I. and Hutter, F. (2017). SGDR: Stochastic gradient descent with warm restarts. In McCallum, A., Bok, M., and Spector, M., editors, *International Conference on Learning Representations*.
- Manning, C. D., Raghavan, P., and Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.
- Martin, F., Lohrmann, G., and Stoll, A. (2022). Selective weed control in grassland using high-pressure water jets. In VDI Wissensforum GmbH, editor, *LAND.TECHNIK 2022*, VDI-Berichte, pages 105–110.
- Neubeck, A. and Van Gool, L. (2006). Efficient non-maximum suppression. In *18th International Conference on Pattern Recognition (ICPR’06)*, pages 850–855. IEEE.
- Petrich, L., Lohrmann, G., Neumann, M., Martin, F., Frey, A., Stoll, A., and Schmidt, V. (2020). Detection of Colchicum autumnale in drone images, using a machine-learning approach. *Precision Agriculture*, 21:1291–1303.
- Schneider, G. R., Scharinger, J., and Probst, C. (2022). Weed detection in grassland and field areas employing RGB imagery with a deep learning algorithm using Rumex obtusifolius plants as a case study. In Falcone, F., Mariani, S., and Laheurte, J.-M., editors, *The 9th International Electronic Conference on Sensors and Applications*, page 87. MDPI.

- Tan, M. and Le, Q. (2021). EfficientNetV2: Smaller models and faster training. In Meila, M. and Zhang, T., editors, *Proceedings of the 38th International Conference on Machine Learning*, pages 10096–10106. PMLR.
- Tan, M., Pang, R., and Le, Q. V. (2020). EfficientDet: Scalable and efficient object detection. In O’Conner, L., editor, *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10778–10787. IEEE.
- Tkachenko, M., Malyuk, M., Holmanyuk, A., and Liubimov, N. (2020). Label Studio: Data labeling software. <https://github.com/HumanSignal/label-studio>. last accessed 2025-03-25.
- Valente, J., Doldersum, M., Roers, C., and Kooistra, L. (2019). Detecting Rumex obtusifolius weed plants in grasslands from UAV RGB imagery using deep learning. In Vosselman, G., Oude Elberink, S. J., and Yang, M. Y., editors, *ISPRS Geospatial Week 2019*, pages 179–185. ISPRS.
- Valente, J., Hiremath, S., Ariza-Sentís, M., Doldersum, M., and Kooistra, L. (2022). Mapping of Rumex obtusifolius in nature conservation areas using very high resolution UAV imagery and deep learning. *International Journal of Applied Earth Observation and Geoinformation*, 112:102864.
- Zauner, C. (2010). *Implementation and Benchmarking of Perceptual Image Hash Functions*. Diplomarbeit, Upper Austria University of Applied Sciences.