

Detection of *Colchicum autumnale* in drone images, using a machine-learning approach

Lukas Petrich^{1,2}, Georg Lohrmann², Matthias Neumann¹, Fabio Martin², Andreas Frey², Albert Stoll², Volker Schmidt¹

¹ Institute of Stochastics, Ulm University, D-89069 Ulm, Germany

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

Abstract

Colchicum autumnale are toxic autumn-blooming flowering plants, which often grow on extensive meadows and pastures. Thus, they pose a threat to farm animals especially in hay and silage. Intensive grassland management or the use of herbicides could reduce these weeds but environment protection requirements often prohibit these measures. For this reason, a non-chemical site- or plant-specific weed control is sought, which aims only at a small area around the *Colchicum autumnale* and with low impact on the surrounding flora and fauna. For this purpose, however, the exact locations of the plants must be known.

In the present paper, a procedure to locate blooming *Colchicum autumnale* in high-resolution drone images in the visible light range is presented. This approach relies on convolutional neural networks to detect the flower positions. The training data, which is based on hand-labeled images, is further enhanced through image augmentation. The quality of the detection was evaluated in particular for grassland sites which were not included in the training to get an estimate for how well the detector works on

previously unseen sites. In this case, 88.6% of the flowers in the test dataset were detected, which makes it suitable, e.g., for applications where the training is performed by the manufacturer of an automatic treatment tool and where the practitioners apply it to their previously unseen grassland sites.

Keywords

Colchicum autumnale; convolutional neural network; drone image; object detection

Corresponding author

Lukas Petrich; Mail: lukas.petrich@uni-ulm.de; Phone: +49 731/50-23590

Introduction

Extensive grassland sites are mowed usually once or twice a year and are not or little fertilized. The typical late mowing dates in the middle of June for these sites help the proliferation of *Colchicum autumnale* on damp and clay soils. A change in the grassland management is often not possible, because of environment protection requirements. It excludes also the use of herbicides. This is a problem for farmers who want to utilize the grassland to feed their animals as all parts of *Colchicum autumnale* are poisonous and can lead to respiratory paralysis and death. In pastures, animals usually avoid the plants. However, leaves and seed in hay and silage cannot be detected by the animals and have to be avoided.

State of the art in control of *Colchicum autumnale* is area-wide mulching of the overall site. The studies of Seither and Elsässer (2014) showed that the above ground population could be reduced by mulching in April or May, at the time of maximum depleted reserve of the onion. However, area-wide mulching contradicts the required extensive management of this kind of grassland. Furthermore, it also has a negative effect on the growth of the crop. It thus reduces the habitat for insects in spring and leads to less grassland yield in summer. Hay cuts in June or July, which are typical for these sites, did not contribute to the reduction of *Colchicum autumnale*. With regard to the fauna, flail mulchers in particular show the greatest damage (Lösch et al., 1997). Löbbert (2001) evaluated the technical procedures, with special reference to the invertebrate fauna. Especially the mulching processes harmed the invertebrates significantly. High damage rates with rotary mowers were demonstrated by Humbert et al. (2010), who consequently recommend leaving uncut areas as a refuge for invertebrates. Braschler et al. (2009) investigated the effect of frequent mowing on population density and species diversity of orthopterans over 7 years. This

study considers frequent mowing one of the main threats to orthopteran communities.

Thus, there is a conflict between management requirements with late mowing dates, grassland yield and fauna protection when *Colchicum autumnale* has to be controlled on extensive grassland sites. The aim is to solve this conflict by non-chemical site- or plant-specific weed control with low impact on the grassland.

It is expected that the leaves of *Colchicum autumnale* will be difficult to distinguish in green grassland in spring with a computer vision system. This problem can be addressed by creating an application map for an automated treatment tool in advance based on image data from a period when *Colchicum autumnale* can be easily distinguished. A further advantage of such a procedure is that areas with no or low density of *Colchicum autumnale* can be identified before, and must not be passed by, the tractor or unmanned vehicle. Resources like worktime and energy can be saved and machine traffic on the grassland is reduced compared to an area-wide online inspection with a ground vehicle like a tractor. *Colchicum autumnale* shows purple flowers in autumn, which differ clearly from the surrounding grassland. Given a procedure to locate each *Colchicum autumnale* on the basis of the flowers, an application map can be created. This map is subsequently used in spring when the plants can be treated effectively with non-chemical tools.

To the best of the authors' knowledge, no previous articles about the automated detection of *Colchicum autumnale* have been published. When it comes to identifying individual plants on grassland, however, a lot of work has been done for the location of *Rumex obtusifolius*. For a review see, e.g., Binch and Fox (2017). While most of those techniques rely on close range images, there are also first attempts to apply deep learning to classify drone images as to whether they show the plant or not (Valente et al., 2019). But

the direct location of *Rumex obtusifolius* in the images was not discussed in the latter study. By processing possibly overlapping cutouts of a larger image individually (sliding window), it is still possible to create maps which show where the plants are located in the original image. However, in order to achieve a high resolution of these maps, it is necessary to evaluate the same areas of the input image several times, which makes direct location techniques, which process the input images only once, more efficient.

The main objective of the present paper is to develop a method for the detection of *Colchicum autumnale* flowers with the help of drone images from a standard RGB camera carried by a multicopter. Moreover, it should perform well in the field under real world conditions.

Materials and Methods

The presented detector solved this problem using a machine learning approach and instead of predicting only the locations, an image segmentation technique was applied to identify pixels in the drone images which belonged to the flowers. The reason for this is that the detector had to be able to handle large drone images coming from different cameras with different resolutions. To deal with this, the drone images were divided into smaller image tiles with a consistent size, on which the flowers were then detected by a neural network. Since the results were maps with values in $[0,1]$ and could thus be interpreted as probabilities that a pixel belonged to a flower, it was necessary to apply a threshold to arrive at binary segmentation masks which classified each input pixel of the image tiles to either being part of a *Colchicum autumnale* flower or of the background. It was then very easy to recombine these masks to end up with segmentation masks of the same size as the original drone image. This recombination would have been much harder if the detection had been based on points or bounding boxes because, in this case, boundary effects like flowers which

were cut in half would have made the recombination much more error-prone. The individual steps of the detector are illustrated in Figure 1.

In order to evaluate how well the proposed detector performed, two different testing regimes were employed. Under the first one, the test images were selected at random from the ground truth dataset. This resembles the standard testing approach in machine learning. Under the second one, for each grassland site a separate model was trained on the images from the remaining grassland sites, and used the selected one only for testing. The purpose of this regime is to evaluate the scenario where a detection model is trained by the manufacturer of an automated treatment tool and then applied by practitioners on their grasslands that are unknown to the manufacturer. Without the need to acquire additional labeled images of their specific fields and fine-tune the model, the treatment tool is much easier to adopt.

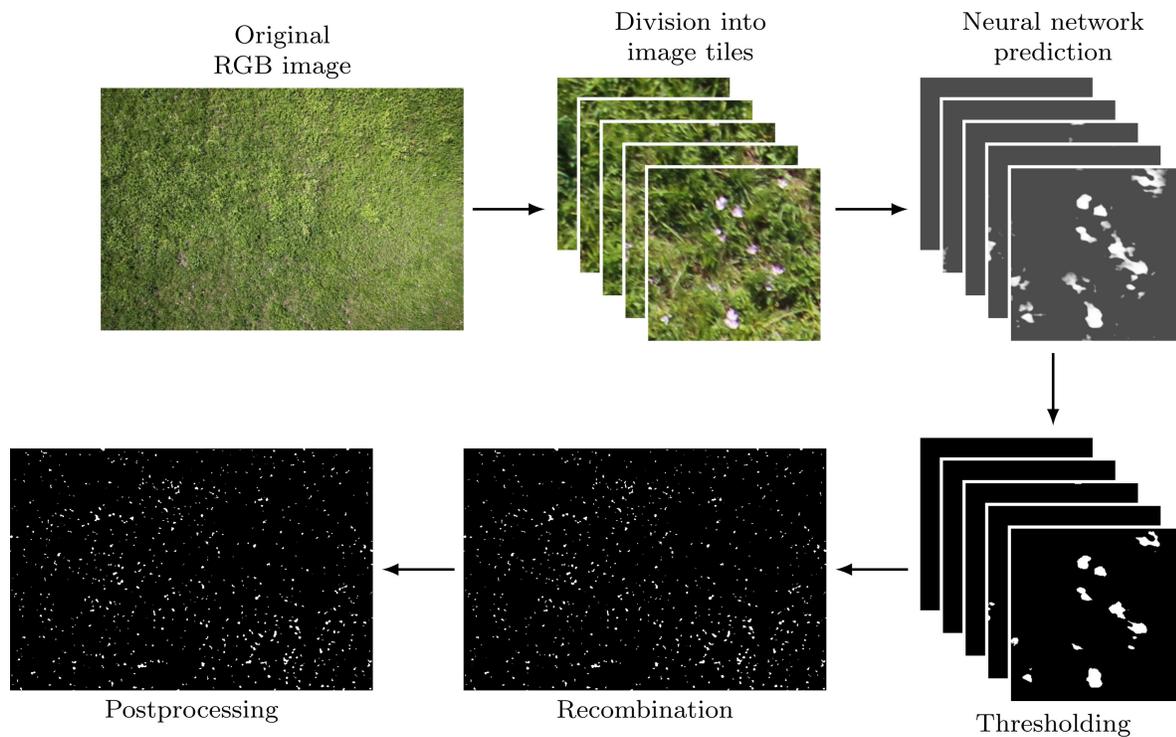


Figure 1: Overview of the individual steps of the *Colchicum autumnale* detector.

Data Description

The images used in this investigation were taken in the period from August 21st to October 1st, 2018. The extensive grassland fields with a size of 1000 to 4000 m² were located in different regions in Baden-Württemberg/Southern Germany. The vegetation was approximately 100 mm high with purple *Colchicum autumnale* flowers. In total, three different sites (near Konstanz, Beuren and Nürtingen) provided by LEVKN (Landscape Conservation Association Constance County) were photographed with a Sony (Tokyo, Japan) alpha 7 RII. The camera has a CMOS full-frame image sensor with 42.4 MP. The focal length of the lens was 24 mm. This camera was mounted on a HiSystems (Moormerland, Germany) MK ARF-OktoXL 4S12 octocopter and triggered by the copter. The payload was 2.5 kg and the maximum take-off weight was 5 kg. During the

flight, the camera aligned itself vertically to the ground with a 2-axis gimbal. The copter was equipped with a lithium polymer accumulator having a capacity of 6600 mAh, which allows an operation time of approximately 10 to 15 min. The route planning was performed with a specific software: The area of interest is defined on the map and based on the given camera properties and the desired image overlap, the software calculates the flight route and the trigger points of the camera automatically. Finally, the flight route is transferred to the copter by a telemetry link. The images used in the present paper were taken at 10 m height over ground.

From the given drone images, 56 of them—the maximum number of images that could be labeled in a reasonable time—were selected in such a way that they did not overlap. It was thus impossible that an image used for training depicted the same area of the grassland sites as an image used for validation or testing, and vice versa, which would have nullified the test results. All *Colchicum autumnale* flowers were then labeled by visual inspection. In order to keep the workload feasible, bounding boxes were used instead of more precise pixel-based annotations, which would have been more work intensive to create. In total, the ground truth dataset comprised 8100 *Colchicum autumnale* flowers marked with bounding boxes. To evaluate how the presented procedure would generalize to previously unseen situations, such as different grassland sites, several different splits into training, validation and test datasets were performed, which are described below, in the section entitled Results.

Note that areas of the drone images which were clearly not and never would be part of the grassland were manually excluded. These included roads, creeks and houses, but not inconstant objects like fences or trees which were also present in some drone images. The argument for this procedure was that identifying these areas is a process that a practitioner

has to do only once and can be reused in the following years without alterations. Furthermore, defining these areas might be necessary for creating an application map of an automatic weed control tool anyway.

Because the presented method is an image segmentation technique, it was necessary to process the ground truth dataset to get segmentation masks for the training of the detector. For that, as a first step, the segmentation mask of a given drone image was completely colored in black (background). After that each annotated bounding box in the drone image was considered and each pixel within the bounding box that did not have a green color was assigned to the foreground (white color). More precisely, a pixel was considered to be green if the hue value in the HSV (hue, saturation, value) color space (see, e.g., Hughes et al. (2013)) was between 60° and 180° . If, under this condition, a bounding box was still colored completely in black, the whole bounding box was filled with white to ensure that no flowers were dropped. This process was designed to only remove pixels which are obviously not part of a flower and to let the neural network learn more sophisticated rules which pixels actually are flower. In summary, the ground truth dataset consisted of 56 drone images together with their corresponding annotations in the form of bounding boxes, which were mostly intended for the evaluation of the detector, and in the form of segmentation masks, which were mostly intended to train the neural network. Note that the dataset splits determined which subsets of the ground truth dataset were actually used for training and evaluation.

Flower Detector

Since the neural network required consistent image sizes, the drone images and their corresponding segmentation masks were divided into smaller 256×256 -pixel tiles. In cases where the drone image dimensions were not multiples of 256, black pixels were used as padding.

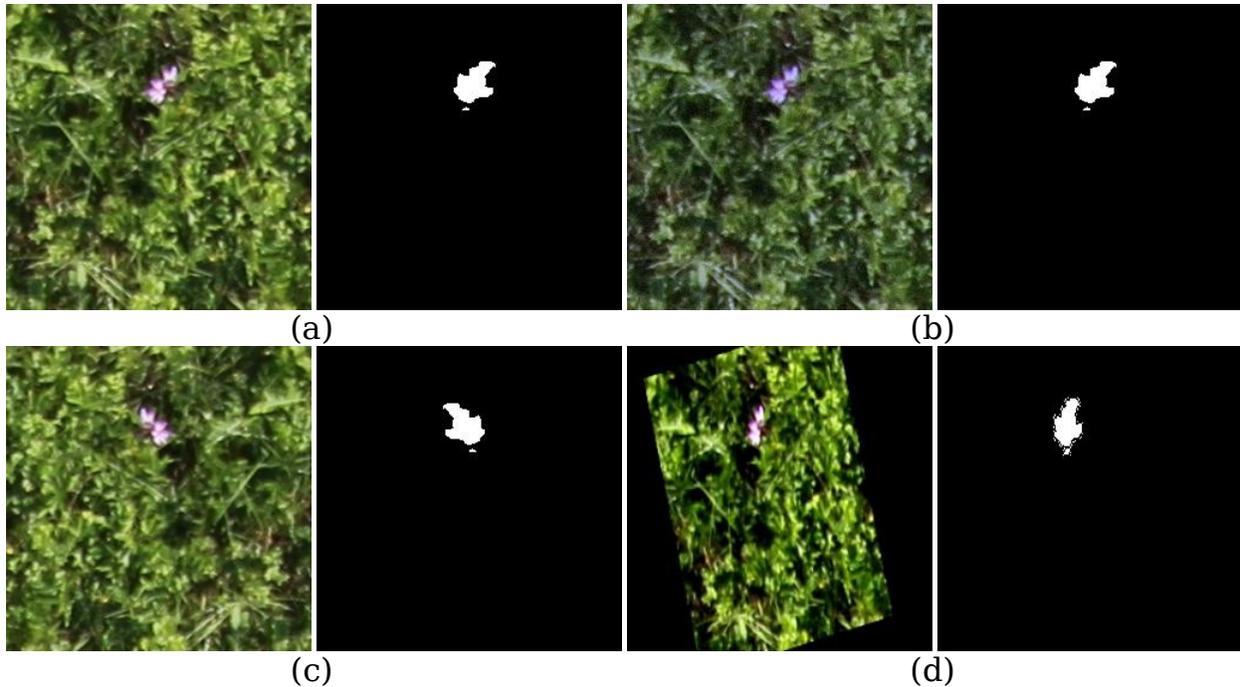


Figure 2: Original input image tile with its corresponding segmentation mask (a) and three different augmentations (b-d).

Image augmentation was employed in order to artificially enlarge the available training dataset for an improvement of the performance of the neural network, see, e.g., Goodfellow et al. (2017). For this, each input image tile and corresponding segmentation mask were subjected to random transformations—i.e., flipping, cropping, Gaussian blurring, contrast adjustments, additive Gaussian noise, changes in the brightness and affine transformations each with random parameters. It was decided at random which of these transformations were applied to each sample. Examples of this process can be seen in Figure 2.

The task of locating the *Colchicum autumnale* can be formulated as a semantic segmentation problem. That is, for each pixel of a given input image, it is predicted whether it shows a flower (foreground) or something else (background). In the present paper, this problem was solved by means of a deep convolutional neural network, see, e.g., Goodfellow et al. (2017)

for more information about this topic. More precisely, the employed architecture was a modification of the U-Net architecture introduced in Ronneberger et al. (2015).

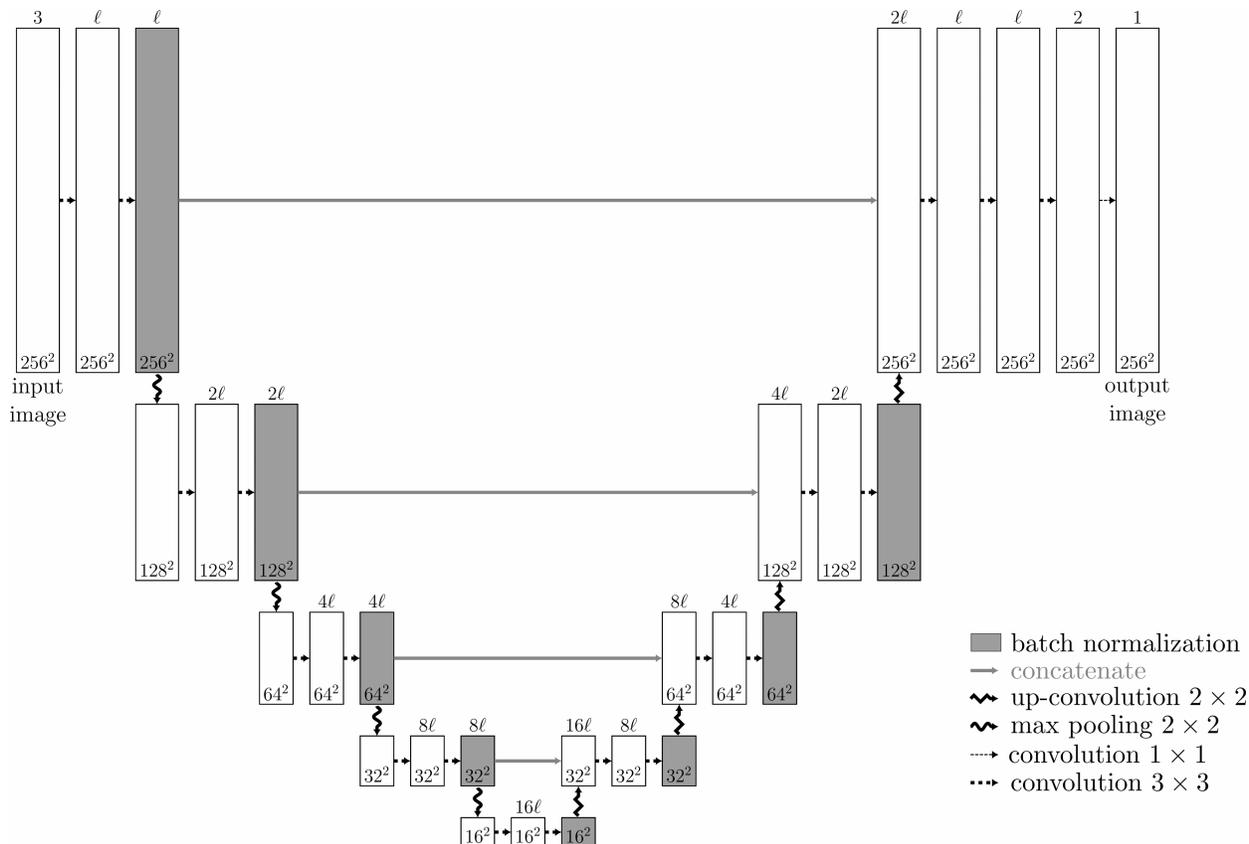


Figure 3: Architecture of the neural network which takes 256×256 RGB image tiles and predicts for each pixel whether it is part of a *Colchicum autumnale* flower or of the background.

In Figure 3 an illustration of the architecture of the neural network can be seen. It could be split into a contracting part and an expansive part. The first one consisted of repeating two 3×3 convolutional layers (with zero-padded boundaries and rectified linear unit (ReLU) activation functions), batch normalization (Goodfellow et al., 2017) and max-pooling layers (with stride 2). In each of the four repetitions, the sizes of the feature maps were cut in half while the number of filters was doubled. The number of filters in the first convolutional layers l_c was determined by hyper-parameter tuning.

In the second part, the contracting part, the reverse happened and the feature maps were enlarged to the same size as the input image. For this, the output of an up-convolution layer, i.e., upsampling combined with a 2×2 convolutional layer, was concatenated with the corresponding feature map from the contracting part and subsequently inserted into two 3×3 convolutional layers (with zero-padded boundaries and ReLU activation functions) followed by a batch normalization layer. These steps were repeated four times. To reduce the number of filters to the number of classes (that is two, one for the foreground and one for the background), a further convolutional layer was added, which was followed by a 1×1 convolutional layer (with sigmoid activation functions). The output was a 256×256 image with one channel where each pixel value is in $[0,1]$. Here, values close to 1 signified *Colchicum autumnale* flowers, while values close to 0 corresponded to background pixels. The neural network was implemented in Tensorflow, see Abadi et al. (2015).

The pixels of the output images of the neural network had continuous values in $[0,1]$ and had to be post-processed in order to end up with binary predictions. For this, a global decision threshold $t_p \in [0,1]$ was applied as binarization. Furthermore, morphological closing (see, e.g., Burger and Burge (2016)) with a disk of radius $r_c > 0$ aggregates clusters of white pixels, and clusters which were still too small to be flowers, i.e., those comprised of fewer than k_s pixels, were removed in the final step.

Training

Consider a split of the ground truth dataset into training, validation and test datasets. For each hyper-parameter configuration, a model of the detector was trained relying only on the training dataset. The best model was selected based on the validation dataset and its performance was evaluated on the test dataset. In the following, the training and the model selection

will be described in more detail.

To train the weights of the neural network, it is important to measure the quality of each prediction. This was done by considering each pixel and assigning a positive loss value if the prediction did not match the corresponding pixel type in the ground truth segmentation masks, and a loss value equal to zero otherwise. More precisely, for the i -th pixel with $i=1, \dots, M$ and M the number of pixels used for training, denote the ground truth pixel label by $y^{(i)} \in \{0,1\}$ and the corresponding prediction of the neural network by $p^{(i)} \in [0,1]$. Moreover, to simplify the notation let

$$\tilde{p}^{(i)} = \tilde{p}^{(i)}(y^{(i)}) = \begin{cases} p^{(i)}, & \text{if } y^{(i)}=1, \\ 1-p^{(i)}, & \text{if } y^{(i)}=0. \end{cases} \quad (1)$$

Obviously, a good prediction aims to always keep $\tilde{p}^{(i)}$ close to 1, and thus loss functions penalize small values of $\tilde{p}^{(i)}$. With this, the classical cross-entropy loss (CE) is given by

$$l^{(CE)}(\tilde{p}^{(i)}) = -\log(\tilde{p}^{(i)}) = \begin{cases} -\log(p^{(i)}), & \text{if } y^{(i)}=1, \\ -\log(1-p^{(i)}), & \text{if } y^{(i)}=0, \end{cases} \quad (2)$$

see, e.g., Goodfellow et al. (2017).

One challenge when training the proposed neural network is that the number of background pixels exceeded the number of foreground pixels by far. The reason for this is that the trivial prediction which assigns background to all pixels already achieves a very small cross-entropy loss value and the features of the *Colchicum autumnale* flowers are thus not learned. This can be solved by introducing a weighting factor $\alpha > 0$ and defining the α -weighted cross-entropy loss, see e.g. Ronneberger et al. (2015), as

$$l_{\alpha}^{(CE)}(\tilde{p}^{(i)}) = -\tilde{\alpha} \log(\tilde{p}^{(i)}) \quad (3)$$

where $\tilde{\alpha} = \alpha 1_{y^{(i)}=1} + (1 - \alpha) 1_{y^{(i)}=0}$. The weight α is usually either chosen through hyper-parameter tuning or directly based on the training dataset. In the present paper, it was set to

$$\alpha = \frac{n_p^{-1}}{n_p^{-1} + n_n^{-1}} \quad (4)$$

with

$$n_p = \#\{i \in \{1, \dots, M\} : y^{(i)} = 1\} \text{ and } n_n = \#\{i \in \{1, \dots, M\} : y^{(i)} = 0\}. \quad (5)$$

Another option that was investigated is a generalization of the cross-entropy loss, the focal loss (FL) introduced in Lin et al. (2017). For the focusing parameter $\gamma \geq 0$, the α -balanced focal loss is given by

$$l_\alpha^{(FL)}(\tilde{p}^{(i)}) = -\tilde{\alpha} (1 - \tilde{p}^{(i)})^\gamma \log(\tilde{p}^{(i)}) \quad (6)$$

It can easily be seen that for $\gamma = 0$ the focal loss is equal to the cross-entropy loss. However, for $\gamma > 0$ the focal loss introduces a multiplicative weight, which reduces the loss for easily classifiable pixels, i.e., where $\tilde{p}^{(i)}$ is vanishing. The training thus concentrates on the misclassified samples, which are downscaled far less. The value $\gamma = 2$, which is suggested in Lin et al. (2017), was used in the experiments presented. The best loss function was determined through hyper-parameter optimization.

The weights of the neural network were trained with the Adam optimization method (Goodfellow et al., 2017) with a batch size of 32 and the learning rate $\lambda_r = 0.0001$, which determines to what extent the newly acquired gradient information in an iteration of the optimization overrides the old information. In order to reduce the runtime, only samples with foreground pixels were used. The training was stopped after 100 optimization iterations through the whole training dataset (so-called epochs) or three days of runtime on an Intel Xeon E5-2670, whichever happened first.

Consider the binary classification problem of assigning each pixel to the

correct class (foreground or background). For the number of true positives n_{tp} , false positives n_{fp} , and false negatives n_{fn} , the precision, recall and F_β -score are given by

$$s_{\text{prec}} = \frac{n_{tp}}{n_{tp} + n_{fp}}, s_{\text{rec}} = \frac{n_{tp}}{n_{tp} + n_{fn}} \text{ and } s_{F_\beta} = \frac{(\beta^2 + 1) s_{\text{prec}} s_{\text{rec}}}{\beta^2 s_{\text{prec}} + s_{\text{rec}}} \quad (7)$$

respectively, where $\beta \geq 0$ is some weighting parameter. Note that the F_β -score is a generalization of the well-known F_1 -score, where by s_{F_β} precision and recall are weighted differently depending on the value of β . For the purposes of weed control, it is beneficial to put a higher emphasis on the recall since it is better to have a higher number of false positives than to miss any *Colchicum autumnale*. Hence, the F_2 -score was used in the experiments. In cases where this assertion does not hold, other values of β can be used. For more information on performance metrics for classification problems see, e.g., Goodfellow et al. (2017) and Manning et al. (2008). With this in mind, the parameters for the postprocessing were chosen based on the training data as follows. The decision threshold t_p , as defined above, was determined by applying different threshold values and maximizing the resulting F_2 -score. Because it was assumed that the shapes of the flowers did not exhibit anisotropy, the radius r_c for the morphological closing was set to the mean length of both sides of all bounding boxes. Finally, the threshold k_s , which identifies small foreground clusters, was determined by the 0.1%-quantile of the volumes of all bounding boxes.

Model Selection

From a practitioner's point of view, the crucial question is not about how precise the predicted segmentation masks are, but rather how well the individual *Colchicum autumnale* flowers are found. For this reason, the binary predictions were also analyzed using a cluster-based approach. For that, each annotated bounding box of a drone image was considered and it

was checked whether there was a foreground cluster in the predicted segmentation mask that had a non-empty intersection with that bounding box. In this case, it was marked as true positive. If there was no corresponding foreground cluster, it was characterized as false negative. After that, every foreground cluster which could not be assigned to a bounding box was marked as false positive. Note that with this procedure, it is not possible to get true negatives and that a foreground cluster can be associated with more than one bounding box. Based on this data, it was possible to compute the precision, the recall and the F_2 -score to evaluate how well the detection models were able to predict the flowers.

Recall that the considered hyper-parameters were the base number of convolutional filters $l_c \in \{8,16,32\}$, and which loss function, either the cross-entropy or the focal loss, should be utilized for the neural network. For each possible configuration of hyper-parameters, a model was trained on the training data and the best values were then determined by maximizing the F_2 -score of the cluster-based evaluation on the validation dataset.

Results

In this section, the test results of the best trained detection models are presented. For that, different splits of the ground truth dataset were used. First, all grassland sites were considered and each drone image was assigned at random to either be part of the training, the validation or the test dataset. The intention of this dataset split is to do an in-depth analysis for the use case where there are training labels for all grassland sites. On the other hand, one might be interested in how well the predictive quality of the detector is for grassland sites which are not part of the training and validation dataset. That is why three further splits of the ground truth dataset were considered where in each one grassland site was selected that is solely used for testing. The two remaining sites formed the training and

validation datasets. Because the model selection was performed in the same way for each of the four dataset splits, all of them were independent of each other and there was no information leakage.

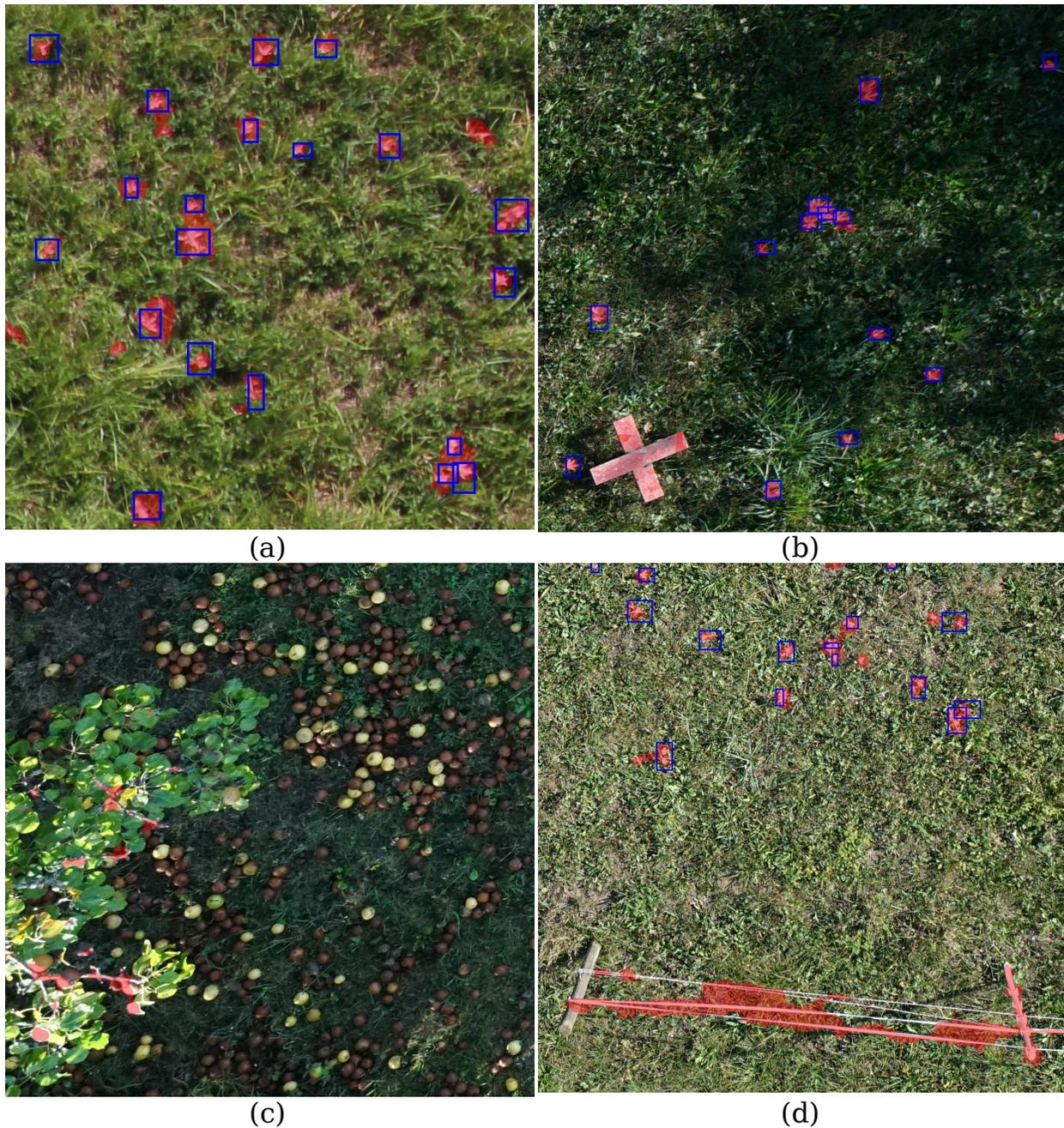


Figure 4: Cutouts of the drone images from the test dataset of the random dataset split overlaid with the predicted segmentation masks of the detection model and the ground truth bounding boxes. On the grassland, most predictions are correct, even in mixed lighting. However, objects like a marker cross (b), tree branches (c) or fences (d) can lead to false positives.

Random Dataset Split

For the first dataset split of the ground truth data, each drone image was assigned at random to the training (34 images), the validation (8 images) and the test dataset (14 images). The best performing model on the validation set turned out to be the one with $l_c=16$ and the cross-entropy loss function. For the test images, the cluster-based evaluation yielded $n_{tp}=1397$ true positives, $n_{fp}=1048$ false positives, and $n_{fn}=20$ false negatives which, in turn, resulted in a precision of 0.571, a recall of 0.986 and a F_2 -score of 0.861. Example cutouts of the drone images, their annotated bounding boxes and the corresponding predictions of the detection model can be seen in Figure 4.

In grassy areas, the prediction was very good (Figure 4a,d), even in deep shadow (Figure 4b). While some false positives were visible, many of them were the result of interfering objects with a light reddish color like tree branches (Figure 4c) or fences (Figure 4d). Other objects like brown apples (Figure 4c) were not wrongly detected. No influence of the dryness of the grass on the predictive performance could be observed.

Site-specific Dataset Split

	Site 1	Site 2	Site 3	Overall
Precision	0.160	0.845	0.488	0.692
Recall	0.897	0.869	0.965	0.886
F_2 -score	0.468	0.864	0.807	0.839
n_{tp}	140	5651	1341	7132
n_{fn}	16	851	48	915
n_{fp}	733	1033	1408	3174
# test images	16	29	11	-
# training images	35	15	31	-
# test flowers	171	6531	1398	-
# training flowers	7608	749	4907	-
Loss function	CE	FL	CE	-
l_c	16	32	16	-

Table 1: Classification metrics of the predicted foreground clusters for drone images of the given grassland site and their aggregated values together with characteristics of the test and training datasets and the best hyper-parameter configuration. The imbalanced numbers of test flowers of the grassland sites were accounted for by first collecting the results of all sites and then based on these the overall precision, recall and F_2 -score were computed (so-called micro-averaging).

Another important question is how well the detector works for previously unseen grassland sites. In order to investigate this, three detection models were trained only on two of the three grassland sites. The remaining one was used for testing. The results of the evaluation are summarized in Table 1.

Site 1 had, compared to the other sites, a high number of other objects, like fences or trees, on the drone images. As discussed above, these were more likely to produce false positives (cf. Figure 4), which, when combined with the low number of flowers in the dataset (cf. Table 1), led to the low precision. Site 2, on the other hand, was comprised mostly of grassy areas. Here the best precision was observed even though the number of flowers in the training dataset was the lowest. Finally, the third grassland site also contained some interfering objects, but the higher number of flowers meant that the precision was not so sensitive to false positives.

Discussion

The detector showed very good results for pure grassland. With recall values between 0.869 and 0.986, a very large portion of the *Colchicum autumnale* flowers were found. Note that about four out of ten predicted locations of a flower were false positives. Many of these were the result of interfering objects like trees and fences, and it is therefore advised to remove these if possible when taking the drone images.

When analyzing the detection models on previously unseen grassland sites, the recall remained stable on high values. The precision varied much more which was mainly caused by interfering objects. However, when applying the detector under real world conditions, a much larger area is imaged and analyzed, most of which will show only grassy areas and only very few parts contain something else. In this case, interfering objects have relatively little impact on the overall precision. Moreover, detections with a low precision lead to a reduced crop yield when the predicted areas are mulched, but not to a deterioration of the weed control results which is more important in the long run. In cases where a higher precision at cost of a lower recall is desired, using, e.g., the F_1 -score instead of the F_2 -score for the parameter estimation would be possible.

With only one exception, the best hyper-parameter configuration turned out to be $l_c=16$ and the cross-entropy loss function. It is thus recommended to use these values when applying the detector and no separate hyper-parameter tuning must be performed.

The proposed detector is a novel approach for locating *Colchicum autumnale* in drone images of grassland sites. The choice of input data distinguishes it from most detectors for *Rumex obtusifolius* relying on close range images (Binch and Fox, 2017), which cannot easily be translated to *Colchicum autumnale* because it does not have the prominent form of the

Rumex leaves. On the other hand, compared to the method proposed by Valente et al. (2019), which also uses drone images, the detector directly located the flowers in the input image and no sliding window procedure was required, which led to a higher efficiency.

Future Work

As shown by Seither and Elsässer (2014) an effective strategy for reducing the stock of *Colchicum autumnale* is mulching in late spring. It is currently unknown how the locations of the plants develop between the imaging and the mulching. It is therefore a point of our future research how to combine the autumn images with images shot in spring to make more accurate predictions. Solely relying on spring images seems like an unnecessarily hard problem since a high spatial correlation between the locations of the plants in autumn and the locations in spring is suspected and detecting *Colchicum autumnale* in spring only by the form of their leaves can be very error-prone. Furthermore, in co-operation with KULT Kress Umweltschonende Landtechnik GmbH (Vaihingen an der Enz, Germany), an automated treatment tool is developed for which the detected locations of the *Colchicum autumnale* are used to create an application map with the AGROCOM MAP software from CLAAS (Harsewinkel, Germany).

Conclusion

In the present paper, a detector of *Colchicum autumnale* flowers in drone images was presented. For that the input drone images were cut to image tiles with a consistent size. On these a convolutional neural network with subsequent post-processing predicted the locations of the flowers. The quality of the detection was evaluated on known and previously unseen grassland sites. In the latter case, 88.6% of the test flowers were detected.

Acknowledgments

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. Furthermore, the authors acknowledge support by the state of Baden-Württemberg through bwHPC.

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2016). TensorFlow: Large-scale machine learning on heterogeneous systems. Technical report. arXiv:1603.04467 [cs.DC]. .
- 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.
- Braschler, B., Marini, L., Thommen, G. H., and Baur, B. (2009). Effects of small-scale grassland fragmentation and frequent mowing on population density and species diversity of orthopterans: a long-term study. *Ecological Entomology*, 34(3):321–329.
- Burger, W. and Burge, M. J. (2016). *Digital Image Processing*. New York, USA: Springer.
- Goodfellow, I., Bengio, Y., and Courville, A. (2017). *Deep Learning*. Cambridge, USA: MIT Press.
- Hughes, J. F., van Dam, A., McGuire, M., Sklar, D. F., Foley, J. D., Feiner, S. K., et al. (2013). *Computer Graphics: Principles and Practice*. Boston, USA: Addison-Wesley Professional, 3rd edition.
- Humbert, J.-Y., Ghazoul, J., Sauter, G. J., and Walter, T. (2010). Impact of different meadow mowing techniques on field invertebrates. *Journal of Applied Entomology*, 134(7):592–599.
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). Focal loss for

- dense object detection. Technical report. arXiv:1708.02002 [cs.CV].
- Löbber, M. (2001). Landscape care - evaluating technical systems with special consideration of invertebrates. *Landtechnik*, 56(SH1):234-236.
- Lösch, M., Stauß, D., Wandel, H., and Jungbluth, T. (1997). Der Bio-Cutter (The organic cutter). *Landtechnik*, 52(1):18-19.
- Manning, C. D., Raghavan, P., and Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge, UK: Cambridge University Press.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: convolutional networks for biomedical image segmentation. In Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F., (Eds.), *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015*, pages 234-241. New York, USA: Springer.
- Seither, M. and Elsässer, M. (2014). Bekämpfungsstrategien gegen Herbstzeitlose (*Colchicum autumnale*) und deren Auswirkungen auf die botanische Zusammensetzung artenreicher Wiesen (Pest control strategies against meadow saffron (*Colchicum autumnale*) and their effects on botanical composition). In *58. Jahrestagung der Arbeitsgemeinschaft Grünland und Futterbau der Gesellschaft für Pflanzenbauwissenschaften e.V.* (58. annual meeting of the working group grassland and forage production of the Society for Crop Science), pages 61-67. (Freising, Germany): Bayerische Landesanstalt für Landwirtschaft.
- Valente, J., Doldersum, M., Roers, C., and Kooistra, L. (2019). Detecting *Rumex obtusifolius* weed plants in grasslands from UAV RGB imagery using deep learning. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W5:179-185.