

001

# Robust In-Field Unsupervised Leaf Instance

002

## Segmentation

001

002

003 Anonymous ECCV 2024 Submission

003

004 Paper ID #10

004

005 **Abstract.** We present a novel unsupervised method for segmenting leaf  
006 instances in plant images captured in the field, addressing a critical gap in  
007 agricultural image processing. Our method computes the differential en-  
008 tropy of image derivatives to distinguish high and low-resolution regions,  
009 yielding foreground leaves and background elements. We use non-linear  
010 diffusion to mitigate focus inversion issues and optimize the foreground-  
011 background separation. Individual leaf instances from the conveyed fore-  
012 ground are segmented using a graph-based contour completion algorithm.  
013 Our method is evaluated on the FieldPlant dataset, proving its effec-  
014 tiveness in detecting stressed leaves with a 18% mIoU. We have also  
015 tested the predictive power of our method using the generated unsuper-  
016 vised masks to supervise Mask-RCNN on FieldPlant tomato, obtaining  
017 promising results of 35.7%. Finally, we address the relevant problem of  
018 plant datasets. This work lays the foundation for scalable, annotation-  
019 free plant phenotyping and disease detection systems that could acceler-  
020 ate agricultural research and precision agriculture. Our approach opens  
021 new avenues for unsupervised semantic segmentation in complex, natu-  
022 ral environments where manual labelling is unreasonable because of its  
023 complexity and resource cost.

024 

## 1 Introduction

024

025 Visual assessment, essential to crop research for over a century [5], remains widely  
026 used in agriculture. These methods are however time-consuming, while RGB-  
027 based solutions are increasingly being proposed as a cost-effective and accessi-  
028 ble alternative to visual assessment, building on deep learning (DL) approaches  
029 in plant pathology [4–6, 20], breeding [44, 56] and ecophysiology [15, 35]. More  
030 specifically, leaf segmentation [48] is a crucial task for phenotyping [51], leading  
031 to further analyses such as stress and disease identification [36, 66], counting  
032 them [14, 25], measuring their dimension [28, 54]. In computer vision and machine  
033 learning, such tasks are usually performed on images of real life, typically based  
034 on datasets such as Pascal VOC, COCO, and cityscape, using either pixel labels  
035 or image-level labels (weakly supervised semantic segmentation).

036 Applying deep learning to leaf segmentation in agricultural settings can be  
037 unreliable due to high variability of agricultural environments, including genetic  
038 differences, lighting conditions and sensor configuration. Trained on a limited  
039 dataset, models will rely on spurious correlations [13] rather than meaningful

025

026

027

028

029

030

031

032

033

034

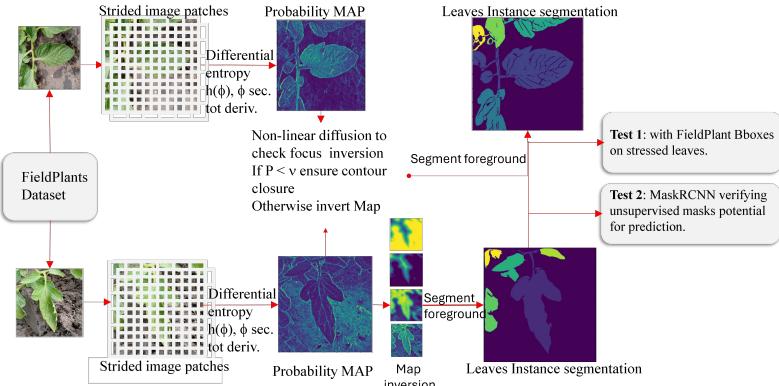
035

036

037

038

039



**Fig. 1:** Model description. Our unsupervised method initially takes care of two cases: the foreground is in focus, or the background is in focus, and which is which is unknown. The figure above shows the main steps leading from foreground-background separation to unsupervised instance segmentation of foreground leaves. Please zoom in.

plant characteristics, resulting in poor generalization performance [24]. Features learned from natural images datasets such as ImageNet have proved inadequate for the nuanced tasks required in plant analysis beyond basic species classification [45]. The complexity of field conditions, with problems such as occlusions and multiple depth planes, further complicates the task and makes foundation models like CLIP [41], STEGO [17] or Segment Anything (SAM) [23] struggle. Adapting foundation models for agriculture will require plant datasets with diverse categorical information required for fine discrimination and pixel-level labelled datasets for field-acquired plant images severely.

These challenges underline the need for specialized, unsupervised approaches capable of adapting to the unique characteristics of agricultural imaging. We propose an unsupervised approach to overcome these limitations and address the unique challenges of field-acquired plant images, such as those in the FieldPlant dataset [33]. The proposed method is better suited to handle the complexities of agricultural environments, including varying depth planes, occlusions, and diverse lighting conditions. By relaxing the constraints of labelling, unsupervised methods can learn more generic features that generalize across non-training domains and avoid the pitfalls of spurious correlations. This approach is precious for plant imaging, where the diversity of environments and the specificity of the task make traditional supervised methods less effective. Our unsupervised method aims to capture the nuanced characteristics of plant images, such as the focus on foreground plants and the separation of multiple view planes, enabling more robust and adaptable plant vision systems for agricultural applications.

Given the unique challenges posed by field conditions, unsupervised instance segmentation of leaves requires a tailored approach. Our contribution is twofold. We provide a novel unsupervised method for separating foreground leaves, those

of interest and thus in focus, and background, namely, any other element irrelevant to the analysis in real complex agricultural environments. Further, given the foreground-background discrimination, we provide, for the first time, unsupervised instance segmentation of plant leaves. This approach overcomes the impossibility of manually labelling all leaves in large datasets, enabling masks to be created for other studies and applications.

## 2 Related Work

**Unsupervised segmentation and Foreground-Background segmentation.** Foreground-background segmentation often assumes different distributions for the foreground and background, for instance assuming a static background for video utilizing PCA or optical flow [7,8]. This idea can be extended and applied to images [46]. However, agricultural datasets pose a challenge due to stationary objects. Alternatives include superpixel segmentation [31] and merging into classes or adapting supervised networks by lowering prediction entropy [55]. Features learned [22] by deep learning (DL) networks and encoder-based segmentation with normalized cut minimization [64] are also used.

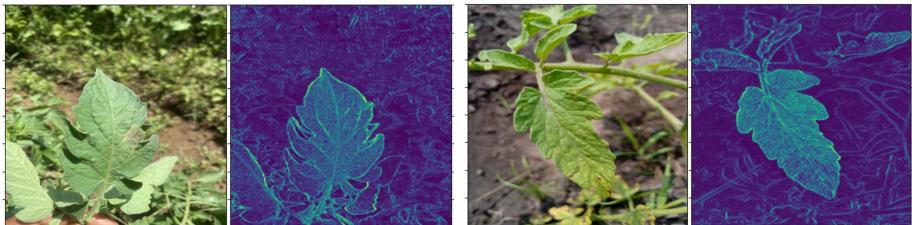
**Application to leaf segmentation.** Leaf segmentation has commonly used color indices [48], non-visible wavelengths like NIR [29], or LAB color space histograms [37]. Recent methods employ deep learning for semantic segmentation [10, 16, 43, 63] but struggle with complex backgrounds such as green background. These approaches can handle more complex backgrounds but are still challenged by similarity in foreground and background vegetation. Unsupervised segmentation is also used in leaf segmentation in agriculture includes using mutual information with VAEs [32] for dynamic mask generation or combining correlation coefficients [21] and deep CNN features for disease recognition. Techniques like super-pixel clustering [67] and K-means for lesion extraction, and AnoLeaf's [3] anomaly localization for disease segmentation are notable. Unsupervised learning is also applied to 3D phenotyping [42] with self-supervised pre-training.

**Datasets.** We focus on leaf segmentation datasets from agricultural species under real-field conditions with plant disease annotations to foster methods aiming to create robust models for addressing practical challenges in agriculture. Instance segmentation of leaves in this context requires extensive, high-quality datasets reflecting various species and conditions in real-field environments, which most datasets often fail to provide due to their controlled backgrounds [11, 19, 30, 47]. In Table 1, we illustrate relevant statistics about some plant datasets. We have not addressed [62] since it treats aerial images taken from a drone.

FieldPlant [33] handles this gap with 5,170 high-resolution images from real fields, showcasing diverse conditions and growth stages of corn, cassava, and tomato leaves, with annotations validated through a process involving agricultural experts. PlantDoc [50], featuring 2,569 images from 13 plant species, offers a multi-crop perspective but exhibits varying image quality due to internet sourc-

**Table 1:** Statistics Dataset

Dataset	Species	Images	Annotations	Anns. per Img.	Classes
FieldPlant [33]	Multi (3)	5,170	8,629	1.67	27
PlantDoc [50]	Multi (13)	2,598	9,216	3.55	30
KaraAgro AI Maize [1]	Maize	15,821	92,934	5.87	8
DiaMOS Plant [12]	Pears	3,505	3,753	1.07	5
Makerere University Beans [34]	Beans	15,402	36,296	2.35	2

**Fig. 2:** Foreground leaves highlighted and well localized by our proposed probability attention map that uses the difference in focus between the foreground and background.

ing. Additionally, there are species-specific datasets [1, 12, 34] with images taken in real fields; their limitation is that they focus on single crops.

### 3 Method

#### 3.1 Rationale and Overview

We consider images of plants taken in the field, for example, by a farmer or a subject interested in analyzing a group of plants in an area, particularly their leaves, possibly from different species. Plants in a field are not isolated but usually occur as a mass of leaves in an image, so distinguishing the leaves is a necessary step for any analysis.

Discriminating the leaves amounts to segmenting them, identifying each one separately, namely performing instance segmentation of leaves against a mass of plants. A first step in instance segmentation is to separate foreground and background. However, relying on pixel supervision, hence manual annotations, would be not only too resource-consuming but perhaps also useless due to the tremendous variety of plants and their leaves. For this reason, we propose unsupervised instance segmentation of foreground leaves (UIS).

The main contributions of our method (see Figure 1) are:

- 1) We provide a method to discriminate foreground plants from background plants and other elements in the background (see, for example, Figure 2).
- 2) We provide unsupervised instance segmentation of foreground leaves (see, for example, Figure 5).
- 3) We show that the segmentation masks that we obtain unsupervisedly have an 18% mIoU with the FieldPlants bounding boxes of stressed leaves, although we have only segmented the foreground leaves (see, for example, Figure 7).

- 133 4) We have used our unsupervised masks as supervision for training a Mask-  
 134 RCNN network, showing some promising generalization results (see Figure 9).  
 135 5) We discuss problems and methodologies related to in-field representativity  
 136 and naturality of plants datasets (see, e.g. Table 4).

137 Our contribution is the very first in this direction. The novelty of our work also  
 138 stems from the fact that we have considered FieldPlant, a dataset of images  
 139 taken in the field under different light conditions, backgrounds, distances, and  
 140 viewpoints.

### 141 3.2 Foreground and background

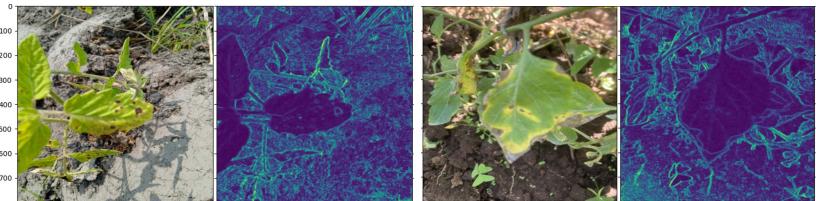
142 The first step for plant image instance segmentation is to separate the fore-  
 143 ground, where the leaves of interest are located, from the background, including  
 144 far leaves, poorly illuminated leaves, and other stuff. Let  $X \in \mathbb{R}^{n \times m \times 3}$  be an  
 145 image from a dataset of images taken in the field. We divide the image into  
 146 square patches of size  $k \times k$  with a stride of  $k - \eta$ , meaning that the patches are  
 147 superimposed by  $\eta$  pixels; the specific measures are detailed in the experiments  
 148 section.

149 We expect that the foreground should be imaged at a higher resolution than  
 150 the background. Let us consider the background as specified by:

$$151 u = Hf + \epsilon \quad (1)$$

152 Where  $f$  is the high-resolution part of the image,  $H = SB$  specifies the blurring  
 153 and down-scaling that affect the background and  $\epsilon$  is the noise. Consider having  
 154 a high-resolution image, and we transform part of it by blurring, downsampling  
 155 and adding noise; then, the image is naturally divided into focused and unfocused  
 156 parts. However, we do not know which part is  $f$  and which part is  $u$ . Let us  
 157 write  $u$  as  $g(f)$ ; we can view the problem in terms of a parameter  $\tau$  to be  
 158 learned such that the image  $X$  is defined as  $X = \tau f + (1 - \tau)g(f)$  which can  
 159 be rewritten as  $\tau = aX + b$ , where  $a = 1/(f - g(f))$  and  $b = -g(f)/(f - g(f))$ ,  
 160 thus obtaining a linear form requiring an equation for each patch leading, for  
 161  $N$  patches, to  $N$  equations and  $3N$  unknowns. This is the way the problem of  
 162 separating foreground and background is treated in the literature, for example  
 163 in [26]. In the matting literature, the optimization of the linear form requires  
 164 some prior input called a trimap [57] or the use of datasets with the provided  
 165 matte parameter [27] to be used with deep networks.

166 In this work, we use the assumption that due to camera focus or, more gen-  
 167 erally, due to attention, the foreground has a higher resolution than the back-  
 168 ground. This prior makes it possible to approximate the solution without manual  
 169 input or labelled datasets. To approximate a solution, we consider  $g(f)$  as ob-  
 170 tained by the high-resolution image  $f$  by applying a Gaussian smoothing filter  
 171 attenuating the image's high frequencies, which implies that the foreground has  
 172 higher differential entropy than the background. We show that differential en-  
 173 tropy is higher in regions with higher resolution, using a simplified representation  
 174 in which the image  $X$  has a single grayscale channel, and each patch is flattened



**Fig. 3:** Inverted probability attention map due to inversion of focus: the background is in focus, while the leaves in the foreground have lower resolution.

175 into a vector with  $x \sim N(\mu, \sigma)$ , with  $\mu$  the patch mean and  $\sigma$  the standard  
176 deviation.

177 Let  $s = 1/(t\sqrt{2\pi}) \exp(-x^2/2t^2)$ , be a Gaussian filter with  $t$  the smoothing  
178 parameter and  $g(f) = f * s$ , with  $*$  the convolution operator. We show that the  
179 differential entropy of the density  $\phi(x) \in \{f, g(f)\}$ :

$$180 \quad h(\phi) = - \int_{\mathcal{X}} \phi \log \phi \text{ in bits} \quad (2) \quad 180$$

181 is higher for  $f$  than for  $g(f)$ . Let  $\phi_{low} = dg(f)/dx$  and  $\phi_{high} = df/dx$ , then the  
182 differential entropy for each is:

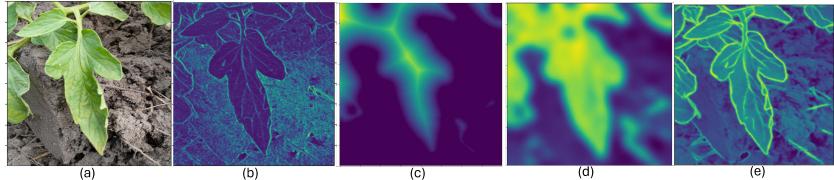
$$183 \quad \begin{aligned} h(\phi_{low}) &= \frac{1}{2(\sigma^2 + t^2)^{3/2}} \sqrt{2\pi} (\sigma^2 1i + t^2 1i) \\ h(\phi_{high}) &= \frac{1}{2\sigma} \sqrt{2\pi} (1i) \end{aligned} \quad (3) \quad 183$$

184 here  $i$  is the imaginary unit. We can see that, even with the same  $\sigma \geq 1$ ,  $h(f) >$   
185  $h(g(f))$ ; moreover, we have to consider that after smoothing the  $\sigma$  related to the  
186 smoothed  $g(f)$ , let it be  $\sigma_{low}$ , is reduced as  $\overline{\sigma_{low}} = \sigma/t^2\sqrt{2\pi}$ .

187 We note that because the Normal has higher entropy than any other distribution,  
188 the simplifying hypothesis does not affect the result. In our model, we actually  
189 consider the second derivatives, using finite difference, in both the  $x$  and the  
190  $y$  directions and the second order total derivative  $f_{xx}(dx)^2 + f_{xy}(dxdy) + f_{yy}(dy)^2$   
191 at each point of a patch  $u \in \mathbb{R}^{k \times k \times 3}$ , disregarding the channel direction, and we  
192 take the square root of its absolute value. Let us indicate the operation  $\|D^2 u\|$   
193 as  $\phi$ , we then take  $h(\phi)$ , namely the differential entropy of  $\phi$ , as indicated in  
194 equation 2, without any need for thresholding and obtain the image  $E_X$ . Examples  
195 of how the differential entropy single out high-resolution parts of the image  
196 are shown in Figure 2. In particular, we denote the outcome map  $E_X$ , separating  
197 the background from the foreground, as the probability attention map.

### 198 3.3 Non Linear Diffusion

199 The image focus or, more generally, the image attention, is not always directed  
200 to the foreground, and examples show that an inversion happens (see figures 3),



**Fig. 4:** Recovering from focus inversion shown in (a) and in the attention map in (b). In (c) the distance transform is applied to the inverted attention map. On the region highlighted by the distance transform non-linear diffusion is applied (d). Finally, the attention map of the original inverted map is obtained (e).

201 creating an analogous inversion on the image  $E_X$ , namely, the background is at  
202 focus. Because the process is unsupervised, it is not possible to know a priori  
203 when the inversion occurs.

204 We propose non-linear diffusion [40] to double-check the inversion. Nonlinear  
205 diffusion filtering was early introduced by Perona and Malik in [39]. The idea is  
206 related to the scale space of images where the diffusion equation acts as a gradient  
207 descent for the minimization of an energy functional  $E[I] = 1/2 \int_{\mathcal{X}} g(\|\nabla I\|^2) dx$ ,  
208 smoothing the image regions respecting the edges. We refer, however, to the  
209 approach of [59–61] introducing coherence-diffusion. Weickert [60] proposes a  
210 structure tensor  $J_t(\nabla u_{\sigma}) = S_t * \nabla u_{\sigma} \nabla u_{\sigma}^T$ , defined by convolving a Gaussian  
211 smoothing filter, with variance  $t$ , with a gradient edge detector  $\nabla u_{\sigma}$  using in turn  
212 a smoother with variance  $\sigma$ . The diffusion tensor  $D(J_t)$  is specified according to a  
213 parameter which depends on the orthogonal eigenvectors  $v_1, v_2$  with eigenvalues  
214  $\mu_1, \mu_2$ , which  $D$  shares with  $J_t$ . We use the diffusion tensor to obtain highly  
215 smoothed regions on  $E_X + Hc$ , where  $Hc$  is the image contour map obtained with  
216 a pretrained holistically nested edge detector [65]. The computation is specified  
217 by the following steps.

218 Let  $\tau(\cdot)$  be the Euclidean distance transform [9], and  $\tau(E_X c + H)$  be the  
219 distance transform of the probability attention map  $E_X$  augmented with the  
220 contour map  $Hc$ , and let us specify the diffusion of the image  $X$  as  $\text{diff}$  then,

$$\begin{aligned} Z_0 &= \text{diff}(X + Hc) \\ Z_t &= \alpha \text{ diff}(Z_{t-1} - \delta_{\mu_t(v)}^{(-)}(Z_{t-1}) + \delta_{\mu_t(v)}^{(+)}(Z_{t-1})) * G_{\sigma} \\ \lambda_t &= (\delta_{\mu_t(v)}^{(-,t)} - \delta_{\mu_t(v)}^{(-,t-1)}, \delta_{\mu_t(v)}^{(+,t)} - \delta_{\mu_t(v)}^{(+,t-1)}) \\ p(E_x) &= \frac{\tau(E_x) - \tau(Z_t)}{\tau(E_x) + \tau(Z_t)} \end{aligned} \quad (4)$$

222 Here,  $\delta_{\mu(v)}^{(\pm)}(Z_t)$  indicates  $Z_t[Z_t \leq \mu(Z_t)] \pm \gamma$ ; namely we subtract a value  $\gamma$ ,  
223 usually 0.1, from the values less than the sample mean  $\mu$  of  $Z_t$  and we add  $\gamma$   
224 to the values greater than the sample mean of  $Z_t$ . In the above displayed (4),  $\alpha$   
225 is a normalization constant.  $G_{\sigma}$  is a low pass filter with parameter  $\sigma$  and  $*$  the  
226 convolution operator.

The tuple  $\lambda$  is used as a stopping criterion when the change in values is under a certain score, yet we have been using  $t \leq 4$ . Finally, if  $p \geq \nu$  then we assume no inversion occurred, otherwise we consider there was an inversion of foreground and background due to a camera or (attention) focus inversion, that is, either far plants or the ground are at focus while close leaves are relatively blurred. We take  $\nu = 0.5$ . If  $p < \nu$  we can, in principle, invert  $E_X$ , however it is better to check the high values in the inverted map, using again the distance transform. An example of the sequence is shown in Figure 4.

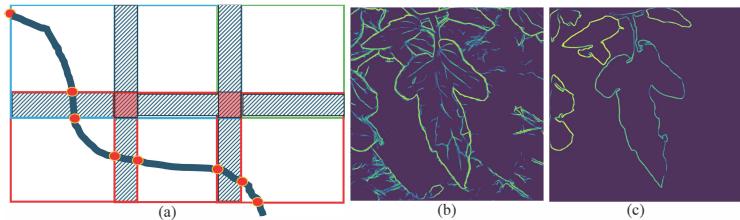


**Fig. 5:** Instance segmentation examples from our method, with good separation of leaves, but with one mixed leaf in the second image from the left.

### 3.4 Instance segmentation

Once the foreground and background have been separated, instance segmentation amounts to label each leaf in the foreground. The first objective is to ensure region consistency; that is, we look for closed contours according to  $E_X$ , namely the probability attention map. Note that, as we are interested in the leaves (actually, there are also the stems), all edges of interest need to be closed to delimit the leaf region. We need to improve the part of the contour obtained by the pretrained holistically nested edge detector to close all open loops pertaining to the foreground.

As for the previous sections, we decompose the images into patches, here with dimension  $k \times k$  with  $k$  large and a stride of dimension  $k - 4$ . We consider a look-ahead strategy that works iteratively. This is defined by a graph in which each node  $n$  (namely a patch) has, as children, its eight neighbouring nodes. When a node (a patch) is visited, the contour lines' extreme points are memorized in a stack together with the index of the successor nodes. A line extreme point inside a patch is where the line ends on the patch boundary. Hence, the successor node



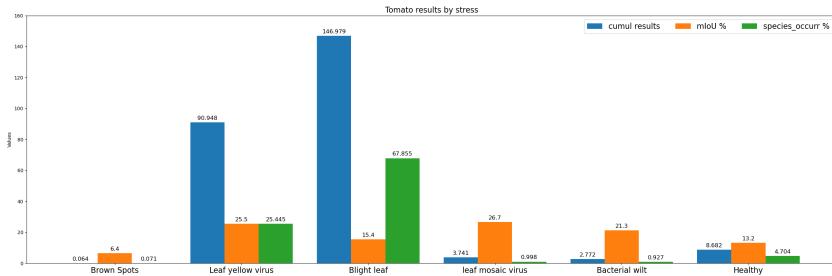
**Fig. 6:** (a) Shows four patches with strides indicated by stripes and the contour intersecting the patches at the red points. The red points form the stack. (b) The contour obtained by the holistically nested edge detector [65]; (c) our final contour.

is the patch adjacent to the boundary, where the line possibly continues. See Figure 6 (a) showing four patches with stride, namely superposition of 4 pixels and the points on the contour intersecting the patch boundaries.

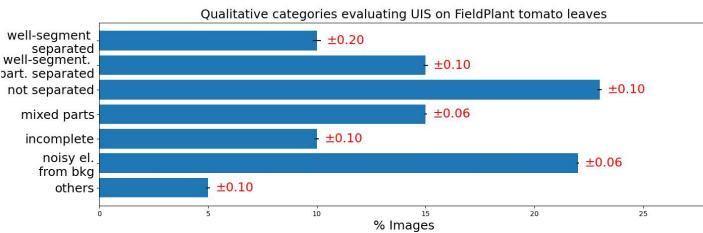
We search the graph in two passes. In the first one, given the sample mean  $\mu_{Hc}$  of the contour matrix  $Hc$ , if the sample mean  $\mu_{current}$  of the current patch is such that  $\mu_{current} > \mu_{Hc}$  then the values higher than  $\beta\mu_{current}$  are considered ( $\beta$  set as 0.97). According to the above control, we collect the location of the line's intersection with the patch boundary, locate headlines not crossing the boundary, and compute the distance with headlines in neighbour patches. A stack of indices of headline locations for each node (patch) is created. All the nodes related to patches with empty boundaries are removed from the stack.

Before the second pass, we use the computed distances of broken heads to connect them to closer line heads (also making multiple connections) and assign them weights according to the distance between the connected heads. We remove all edges with weight less than  $(1/2) \max_{Hc} - \mu_{Hc}$ . Note that the graph's vertices are now located on the patch boundaries; see the intersection points shown in red in Figure 6. Having all the points on the patch boundaries in the second pass, we must follow a path minimizing the cost of a closed boundary. The Floyd-Warshall algorithm is used to find all the paths linking the points defined on the patch boundaries. Given a sequence of pixels  $Y = (x(t_0), y(t_0), \theta(t_0)), (x(t_1), y(t_1), \theta(t_1)), \dots, (x(t_m), y(t_m), \theta(t_m))$ , passing through patch boundaries, we choose the line in  $Hc$  that satisfies the closure, that is,  $\min(|x_m - x_1, y_m - y_1|)$  and is admissible, according to [2], minimizing the functional form  $\int_{t_i}^{t_{i+1}} [(dx/dt)^2 + (dy/dt)^2 + (d\theta/dt)^2] dt$ . Note that we just need to verify that the minimization of the functional is satisfied. In Figure 6, we show the transformation of the contour obtained with the holistic method into its refinement obtained with the above succinctly described contour iterative deepening.

Given the closed regions obtained by the refined contour, we label each region and generate the tensor having a mask for each closed contour, namely, leaf instance.



**Fig. 7:** Bar plot of mIoU% related to the percentage of images for each stress in tomato leaves, and the cumulative IoU for all images for each stress, please zoom in.



**Fig. 8:** Bar plot of qualitative categories, defined by inspecting the results of the UIS on FieldPlants dataset foreground tomato leaves, please zoom in.

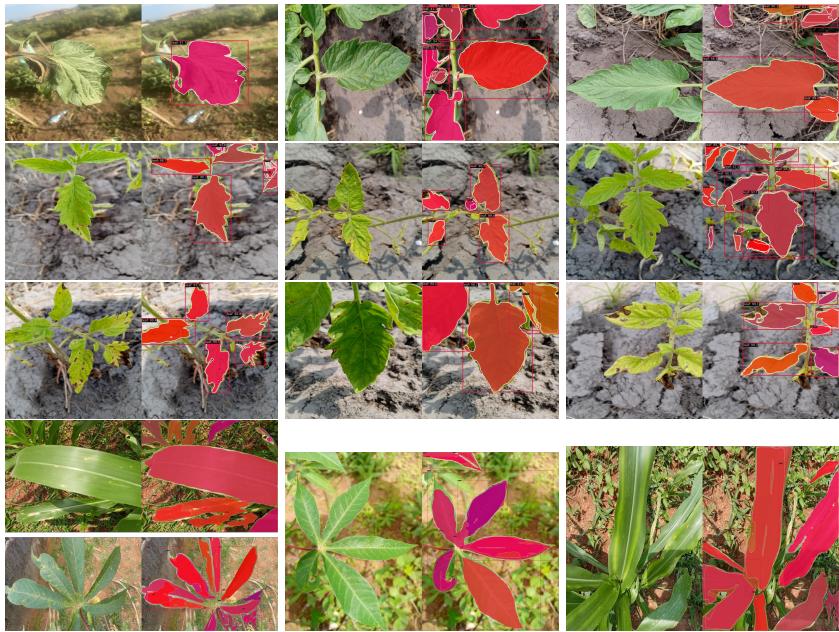
### 282 3.5 Implementation

283 The algorithm is implemented in Python 3.11.7 and Pytorch-cuda 12.1 in a  
 284 Conda environment. The algorithm is run on a GeForce RTX-4090. We initially  
 285 resize all the images in the dataset to size  $800 \times 800 \times 3$ . We transform each image  
 286 into patches of a size that depends on the task. For building the probability  
 287 attention map  $E_X$  described in Subsection 3.2, we use patches of size  $4 \times 4 \times 3$   
 288 with a stride of 2, and similarly in 3.3. For UIS (Subsection 3.4), we use patches  
 289 of size  $16 \times 16$  with a stride of 4. Our method takes an average time of 3 sec. to  
 290 process each image.

291 For testing the method, we have used the Tomato species of the FieldPlant  
 292 dataset with its six stress categories, as illustrated in the Bar-plot 7. The dataset  
 293 is discussed in Section 5.

## 294 4 Results

295 We have experimented with the six stresses of the proposed Unsupervised In-  
 296 stance Segmentation (UIS) model on FieldPlants [33] on the Tomato species.  
 297 The tomato species is challenging for UIS because of aspects concerning resolu-  
 298 tion, focus, colours and, in general, light distribution on the leaves due to their  
 299 glossiness, which, for example, does not affect Cassava or Corn leaves.



**Fig. 9:** Instance segmentation mask prediction with Mask-RCNN using our unsupervised masks for supervision. In the last strip, examples of masks for Cassava and Corn stresses. For Cassava and Corn prediction uses tomato unsupervised masks (unsupervised transfer learning). Zoom in to see leaf prediction confidence.

Not having available ground truth masks, we have considered two tests to compare with other methods despite the fact that no segmentation method is available on FieldPlant, and the UIS method is not available at all on any plant dataset. In the first test, we obtained bounding boxes (bboxes) from our masks to compare the bounding boxes with those available on FieldPlant for the six stresses of tomato plants (see Figure 7).

With the second test, we have trained Mask-RCNN with both our unsupervised masks and bboxes, using as a training set 70% of FieldPlant tomato images and as a test set the remaining 30%, obtaining an average prediction (mAP) of 35.7 and average Recall of 50.1. From the point of view of UIS, not having ground truth, the results are just qualitative; some mask predictions from Mask-RCNN trained with our unsupervised masks, are shown in Figure 9.

**Table 2:** Per stress-class performance mIoU% on the task of leaf stress detection in terms of mIoU% on FieldPlant Tomato species

	Method	Trained on	Tested on	Model	Tot %	Brown Spots	Leaf yellow	Blight leaf	Leaf mosaic	Bacterial wilt	Healthy
Ours	Unsupervised detection	-	FieldPlants	UIS	18.1	6.4	25.5	15.4	26.7	21.3	15.2
Monpojou et al. [33]	Supervised detection	Plant Village	FieldPlant	MobileNet	24.26	-	0	35	0	0	8.9

Note that our objective is to obtain UIS for plant leaves in the foreground, without specifically considering stressed leaves; in particular, we limited our foreground leaves UIS to 5 leaves per image. Because the Field Plants dataset is about stressed plants, a part of the individually segmented leaves is also stressed. To compute the overall mIoU% we have made a bbox around each mask and then computed the IoU% between the FieldPlant bounding boxes and our bboxes, for each image. Finally we averaged the result on the overall number of bounding boxes. Figure 7 gives an overview of the results and the distribution of the 1403 tomato plants images on the different stresses. Given the lack of ground truth, we have defined a qualitative-based evaluation by visually inspecting the results, reported in the bar plot of Figure 8. Here we define 7 categories evaluating the accuracy of the unsupervised masks.

**Table 3:** AP and Recall comparisons on FieldPlant Tomato

	Method	Trained on	Tested on	Model	mAP%	Recall %
Ours	Supervised by UIS masks	FieldPlant Tomato 70%	FieldPlant Tomato %30	Mask-RCnn	35.7	50.1
Ours	Supervised by UIS masks	FieldPlant Tomato 70%	FieldPlant (Cassava, Corn)	Mask-RCnn	20.7	45.8
Moupojou et Al. [33]	Supervised detection	Plant Village	FieldPlant	MobileNet	61.92	-
Moupojou et Al. [33]	Supervised detection	FieldPlant	FieldPlant	MobileNet	82.9	-
Wang et Al. [58]	Supervised detection	FieldPlant	FieldPlant	their method	70.1	69.6

**Comparison with the State of the Art.** We compare our results with the work of [33], and the work of [58]; apparently, there are no other works relying on FieldPlants as also reported in the survey of [38].

To address the comparison we have considered both the tests described above. In particular, concerning the first test, an analysis of the accuracy with respect to stress is provided in [33]. We have considered the confusion matrix in [33] reporting results using MobileNet on tomato leaves stresses to compare with our results on each stress of the tomato species, see Table 2. Note that in the confusion matrix, the authors do not consider the full image but cropped leaves.

Both works [33] and [58] use mAP, which corresponds to COCO  $AP^{IoU=.50}$  and similar for Recall. We use these Mask-RCNN metrics for our results. The results obtained by training Mask-RCNN with UIS masks and results from [33] and [58] are shown in Table 3. Note that on the second row of the table results on the whole FieldPlant dataset are reported based on training with the unsupervised masks just on the tomato species. Despite the gap with supervised detection, these results indicate that we are on the right path to closing the gap with supervised methods.

## 341 5 Discussion on Plant Datasets

342 In this section, we justify the selection of FieldPlant [33] over PlantDoc [50] and  
 343 investigate the consistency of foreground focus in its images.

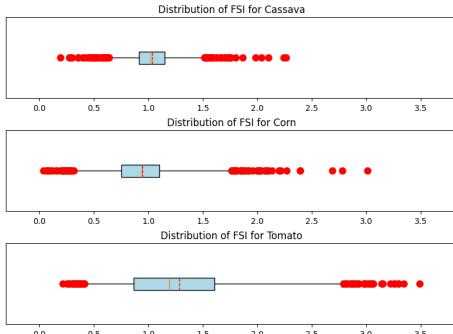
344 **Dataset Comparison for Natural In-Field Contexts Representativity.**  
 345 In [33], the authors conducted a supervised benchmarking study comparing  
 346 FieldPlant with PlantDoc. They evaluated several CNN models [18, 49, 52, 53], on  
 347 both raw and cropped images, showing that FieldPlant’s field-specific conditions  
 348 and larger dataset achieved better classification accuracy than PlantDoc. In a  
 349 complementary analysis, we assessed how well the FieldPlant dataset represents  
 350 natural, in-field conditions compared to PlantDoc by examining several key im-  
 351 age metrics: edge density, color variance, texture complexity, and composition  
 352 complexity. To further validate our findings, we considered the Coffee Leaf Biotic  
 353 Stress dataset (CLBS) [11]. This dataset consists of images of single coffee leaves  
 354 on a consistent background, representing an extremely controlled environment.  
 355 The results are summarized in Table 4.

**Table 4:** Comparison of in-field representativity for FieldPlant, PlantDoc, and Coffee Leaf Biotic Stress. Each cell shows the average followed by the variance in parentheses.

Metric	FieldPlant	PlantDoc	CLBS
Edge Density	0.466 ( $4.000 \times 10^{-4}$ )	0.448 ( $5.800 \times 10^{-3}$ )	0.427 ( $6.900 \times 10^{-4}$ )
Composition Complexity	7.400 (0.069)	7.040 (1.002)	6.880 (0.111)
Color Variance	0.037 ( $1.864 \times 10^{-4}$ )	0.047 ( $5.265 \times 10^{-4}$ )	0.068 ( $1.442 \times 10^{-4}$ )
Texture Complexity	0.011 ( $2.509 \times 10^{-5}$ )	0.014 ( $4.763 \times 10^{-5}$ )	0.021 ( $1.503 \times 10^{-5}$ )

356 Our findings reveal that FieldPlant presents higher edge density and  
 357 composition complexity, indicative of more complex and more intricate real-world  
 358 scenes. In contrast, PlantDoc, which incorporates images from controlled and art-  
 359 ificial contexts, shows slightly lower values for these metrics. Additionally, lower  
 360 color variance and texture complexity in FieldPlant suggest less artificial ma-  
 361 nipulation and a closer representation of natural conditions. Higher values for  
 362 PlantDoc could result from digitally enhanced images or with textures added  
 363 from controlled settings. Higher variance across all metrics for PlantDoc indi-  
 364 cates the presence of images with various non-natural contexts. Consistent with  
 365 its controlled setup, the values of the metrics on CLBS support our approach.

366 **Foreground Sharpness Analysis.** For each species in the FieldPlant dataset—  
 367 cassava, corn, and tomato—we conducted a series of analyses to evaluate the focus  
 368 and sharpness of the images, which are critical for effective leaf segmentation.  
 369 We approximated the foreground using the bounding boxes provided in the  
 370 annotations. For each image, we cropped each region defined by a bounding box  
 371 and calculated the Laplacian variance for these regions; then, we averaged these



**Fig. 10:** Boxplots showing the distribution of the Foreground Sharpness Index (FSI) for cassava, corn, and tomato.

Species	FS	IS	FSI
Cassava	87.84	91.38	1.04
Corn	145.54	165.70	0.95
Tomato	116.16	114.28	1.32

**Table 5:** Average foreground sharpness (FS), average image sharpness (IS), and average Foreground Sharpness Index (FSI) for cassava, corn, and tomato in the FieldPlant dataset.

variances to obtain a foreground sharpness (FS) measure. We also calculated the Laplacian variance of the entire image to represent the overall image sharpness (IS). The Foreground Sharpness Index (FSI) was defined as the ratio of the foreground sharpness to the total image sharpness, providing a comparative measure of the relative focus of the foreground leaves versus the entire image.

Table 5 summarizes the average sharpness and FSI values for each species, while Figure 10 illustrates the distribution of the FSI values across the three species with boxplots. Our analysis reveals that the tomato species exhibits a higher foreground sharpness relative to the overall image sharpness than the other species. Moreover, we observed that some images in the FieldPlant dataset show the background in focus while the foreground is blurred, see for example Figure 3, an effect more pronounced in the corn species as indicated by the FSI values.

## 6 Conclusions

We have presented a novel approach for in-field plant analysis. The novelty of our work lies both in the proposed method and in the principle of dealing with images captured on the field in a variety of outdoor environments with different lighting, backgrounds, distances and viewpoints. In this work, we have used the novel FieldPlant dataset [33] with a view on real agricultural problems.

We have addressed two relevant issues: unsupervised separation of foreground leaves from the background and, based on this, unsupervised instance segmentation of foreground plant leaves. We have shown that our method can generalize to unsupervised stress detection and can be used for training instance segmentation networks, such as Mask-RCNN, based only on unsupervised masks.

We have also addressed some statistical problems related to plant datasets. Despite a gap with supervised methods, our results are promising and open the road to new approaches.

399

## References

399

- 400 1. Akogo, D., Samori, I., Acquaye, C., Addo, M., Amoako, E., Ezroa-Cudjoe, F.,  
401 Buaba, J., Seloame, T.N., Mavis, A., Bright, H., Kezia, G.D., Ababio, Y.O., Clinton,  
402 D.A.: The KaraAgro AI Maize Dataset (2022). <https://doi.org/10.7910/DVN/CXUMDS>, <https://doi.org/10.7910/DVN/CXUMDS> 4  
403
- 404 2. Ben-Yosef, G., Ben-Shahar, O.: Minimum length in the tangent bundle as a model  
405 for curve completion. In: 2010 IEEE Computer Society Conference on Computer  
406 Vision and Pattern Recognition. pp. 2384–2391. IEEE (2010) 9  
407
- 408 3. Bhugra, S., Kaushik, V., Gupta, A., Lall, B., Chaudhury, S.: Anoleaf: Unsupervised  
409 leaf disease segmentation via structurally robust generative inpainting. In: Proceedings  
410 of the IEEE/CVF winter conference on applications of computer vision. pp. 6415–6424 (2023) 3  
411
- 412 4. Bock, C., Barbedo, J., Ponte, E.M., Bohnenkamp, D., Mahlein, A.K.: From visual  
413 estimates to fully automated sensor-based measurements of plant disease severity:  
414 status and challenges for improving accuracy. Phytopathology Research **2**, 1–30  
415 (2020). <https://doi.org/10.1186/s42483-020-00049-8> 1  
416
- 417 5. Bock, C., Chiang, K., Ponte, E.D.D.: Plant disease severity estimated visually:  
418 a century of research, best practices, and opportunities for improving methods  
419 and practices to maximize accuracy. Tropical Plant Pathology **47**, 25 – 42 (2021).  
420 <https://doi.org/10.1007/s40858-021-00439-z> 1  
421
- 422 6. Bock, C., Poole, G., Parker, P., Gottwald, T.: Plant disease severity estimated  
423 visually, by digital photography and image analysis, and by hyperspectral imaging.  
424 Critical Reviews in Plant Sciences **29**, 107 – 59 (2010). <https://doi.org/10.1080/07352681003617285> 1  
425
- 426 7. Bouwmans, T.: Recent Advanced Statistical Background Modeling for Foreground  
427 Detection - A Systematic Survey. Recent Patents on Computer Science **4**(3), 147–  
428 176 (Sep 2011), <https://hal.science/hal-00644746> 3  
429
- 430 8. Bouwmans, T.: Traditional and recent approaches in background modeling for  
431 foreground detection: An overview. Computer science review **11**, 31–66 (2014) 3  
432
- 433 9. Breu, H., Gil, J., Kirkpatrick, D., Werman, M.: Linear time euclidean distance  
434 transform algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence  
435 **17**(5), 529–533 (1995) 7  
436
- 437 10. Champ, J., Mora-Fallas, A., Goëau, H., Mata-Montero, E., Bonnet, P., Joly, A.:  
438 Instance segmentation for the fine detection of crop and weed plants by precision  
439 agricultural robots. Appl Plant Sci **8**(7), e11373 (Jul 2020). <https://doi.org/10.1002/aps3.11373> 3  
440
- 441 11. Esgario, J.G., Krohling, R.A., Ventura, J.A.: Deep learning for classification and  
442 severity estimation of coffee leaf biotic stress. Computers and Electronics in Agriculture  
443 **169**, 105162 (2020) 3, 13  
444
- 445 12. Fenu, G., Mallochi, F.M.: Diamos plant: A dataset for diagnosis and monitoring  
446 plant disease. Agronomy **11**(11) (2021). <https://doi.org/10.3390/agronomy11112107>, <https://www.mdpi.com/2073-4395/11/11/2107> 4  
447
- 448 13. Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F.A., Brendel, W.:  
449 Imagenet-trained cnns are biased towards texture; increasing shape bias improves  
450 accuracy and robustness. CORR **abs/1811.12231** (2018), <http://arxiv.org/abs/1811.12231> 1  
451
- 452 14. Giuffrida, M.V., Minervini, M., Tsaftaris, S.A.: Learning to count leaves in rosette  
453 plants (2016) 1  
454

15. Grokkinsky, D.K., Svensgaard, J., Christensen, S., Roitsch, T.: Plant phenomics and the need for physiological phenotyping across scales to narrow the genotype-to-phenotype knowledge gap. *Journal of Experimental Botany* **66**(18), 5429–5440 (07 2015). <https://doi.org/10.1093/jxb/erv345>, <https://doi.org/10.1093/jxb/erv345> 1
  16. Guo, R., Qu, L., Niu, D., Li, Z., Yue, J.: Leafmask: Towards greater accuracy on leaf segmentation. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 1249–1258 (2021) 3
  17. Hamilton, M., Zhang, Z., Hariharan, B., Snavely, N., Freeman, W.T.: Unsupervised semantic segmentation by distilling feature correspondences (2022), <https://arxiv.org/abs/2203.08414> 2
  18. Howard, A.G., Zhu, M.S., Chen, L., et al.: Mobilenets: Efficient convolutional neural networks for mobile vision applications. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* pp. 2280–2288 (2017). <https://doi.org/10.1109/CVPR.2017.270> 13
  19. Hughes, D.P., Salathe, M.: An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv preprint arXiv:1511.08060* (2015) 3
  20. Jin, D., Yin, H., Zheng, R., Yoo, S.J., Gu, Y.H.: Plantinfocms: Scalable plant disease information collection and management system for training ai models. *Sensors* **23**(11) (2023). <https://doi.org/10.3390/s23115032>, <https://www.mdpi.com/1424-8220/23/11/5032> 1
  21. Khan, M.A., Akram, T., Sharif, M., Awais, M., Javed, K., Ali, H., Saba, T.: Ccdf: Automatic system for segmentation and recognition of fruit crops diseases based on correlation coefficient and deep cnn features. *Computers and Electronics in Agriculture* **155**, 220–236 (2018). <https://doi.org/https://doi.org/10.1016/j.compag.2018.10.013>, <https://www.sciencedirect.com/science/article/pii/S0168169918303120> 3
  22. Kim, W., Kaneko, A., Tanaka, M.: Unsupervised learning of image segmentation based on differentiable feature clustering. *CoRR abs/2007.09990* (2020), <https://arxiv.org/abs/2007.09990> 3
  23. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollár, P., Girshick, R.: Segment anything (2023) 2
  24. Koh, P.W., Sagawa, S., Marklund, H., Xie, S.M., Zhang, M., Balsubramani, A., Hu, W., Yasunaga, M., Phillips, R.L., Gao, I., et al.: Wilds: A benchmark of in-the-wild distribution shifts. In: *International conference on machine learning*. pp. 5637–5664. *PMLR* (2021) 2
  25. Kumar, J.P., Dominic, S.: Image based leaf segmentation and counting in rosette plants. *Information processing in agriculture* **6**(2), 233–246 (2019) 1
  26. Levin, A., Lischinski, D., Weiss, Y.: A closed-form solution to natural image matting. *IEEE transactions on pattern analysis and machine intelligence* **30**(2), 228–242 (2007) 5
  27. Li, J., Jain, J., Shi, H.: Matting anything. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 1775–1785 (2024) 5
  28. Machefer, M., Lemarchand, F., Bonnefond, V., Hitchins, A., Sidiropoulos, P.: Mask r-cnn refitting strategy for plant counting and sizing in uav imagery. *Remote Sensing* **12**(18), 3015 (2020) 1
  29. Milioti, A., Lottes, P., Stachniss, C.: 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*

- 498      **IV-2/W3**, 41–48 (2017). <https://doi.org/10.5194/isprs-annals-IV-2-W3-41-2017>, <https://isprs-annals.copernicus.org/articles/IV-2-W3/41/2017/> 3
- 499
- 500
- 501      30. Minervini, M., Abdelsamea, M.M., Tsafaris, S.A.: Image-based plant phenotyping  
502      with incremental learning and active contours. Ecological Informatics **23**, 35–48  
503      (2014). <https://doi.org/10.1016/j.ecoinf.2013.07.004>,  
504      <https://www.sciencedirect.com/science/article/pii/S1574954113000691>,  
505      special Issue on Multimedia in Ecology and Environment 3
- 506      31. Mirsadeghi, S.E., Royat, A., Rezatofighi, H.: Unsupervised image segmenta-  
507      tion by mutual information maximization and adversarial regularization. CoRR  
508      [abs/2107.00691](https://arxiv.org/abs/2107.00691) (2021), <https://arxiv.org/abs/2107.00691> 3
- 509      32. Mirsadeghi, S.E., Royat, A., Rezatofighi, H.: Unsupervised image segmentation by  
510      mutual information maximization and adversarial regularization (2021), <https://arxiv.org/abs/2107.00691> 3
- 511
- 512      33. Moupojou, E., Tagne, A., Retraint, F., Tadonkemwa, A., Wilfried, D., Tapamo,  
513      H., Nkenlifack, M.: Fieldplant: A dataset of field plant images for plant disease de-  
514      tection and classification with deep learning. IEEE Access **11**, 35398–35410 (2023).  
515      <https://doi.org/10.1109/ACCESS.2023.3263042> 2, 3, 4, 10, 11, 12, 13, 14
- 516
- 517      34. Mugalu, B.W., Nakatumba-Nabende, J., Katumba, A., Babirye, C., Tusubira, F.J.,  
518      Mutebi, C., Nsumba, S., Namanya, G.: Makerere University Beans Image Dataset  
519      (2022). <https://doi.org/10.7910/DVN/TCKVEW>, <https://doi.org/10.7910/DVN/TCKVEW> 4
- 520
- 521      35. Mutka, A.M., Fentress, S.J., Sher, J.W., Berry, J.C., Pretz, C., Nusinow, D.A.,  
522      Bart, R.: Quantitative, Image-Based Phenotyping Methods Provide Insight into  
523      Spatial and Temporal Dimensions of Plant Disease . Plant Physiology **172**(2),  
524      650–660 (09 2016). <https://doi.org/10.1104/pp.16.00984>, <https://doi.org/10.1104/pp.16.00984> 1
- 525
- 526      36. Noon, S.K., Amjad, M., Qureshi, M.A., Mannan, A.: Use of deep learning tech-  
527      niques for identification of plant leaf stresses: A review. Sustainable Computing:  
528      Informatics and Systems **28**, 100443 (2020) 1
- 529
- 530      37. Pape, J.M., Klukas, C.: 3-d histogram-based segmentation and leaf detection for  
531      rosette plants. In: Agapito, L., Bronstein, M.M., Rother, C. (eds.) Computer Vision -  
532      ECCV 2014 Workshops. pp. 61–74. Springer International Publishing, Cham  
533      (2015) 3
- 534
- 535      38. Patil, A., et al.: Survey on machine learning techniques for the identification and  
536      classification of plant leaf diseases. In: 2023 Global Conference on Information  
537      Technologies and Communications (GCITC). pp. 1–6. IEEE (2023) 12
- 538
- 539      39. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic (1988) 7
- 540
- 541      40. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion.  
542      IEEE Transactions on pattern analysis and machine intelligence **12**(7), 629–639  
543      (1990) 7
- 544
- 545      41. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G.,  
546      Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I.: Learning transferable  
547      visual models from natural language supervision. In: Meila, M., Zhang, T. (eds.)  
548      Proceedings of the 38th International Conference on Machine Learning. Proceed-  
549      ings of Machine Learning Research, vol. 139, pp. 8748–8763. PMLR (18–24 Jul  
550      2021), <https://proceedings.mlr.press/v139/radford21a.html> 2
- 551
- 552      42. Roggiolani, G., Magistri, F., Guadagnino, T., Behley, J., Stachniss, C.: Unsuper-  
553      vised pre-training for 3d leaf instance segmentation. IEEE Robotics and Automa-  
554      tion Letters **8**(11), 7448–7455 (Nov 2023). <https://doi.org/10.1109/lra.2023.3320018>  
555      3

- 549 43. Roggiolani, G., Sodano, M., Guadagnino, T., Magistri, F., Behley, J., Stachniss,  
550 C.: Hierarchical approach for joint semantic, plant instance, and leaf instance seg-  
551 mentation in the agricultural domain. In: 2023 IEEE International Conference on  
552 Robotics and Automation (ICRA). pp. 9601–9607. IEEE (2023) 551  
553 44. Roth, L., Fossati, D., Krähenbühl, P., Walter, A., Hund, A.: Image-based phenomic  
554 prediction can provide valuable decision support in wheat breeding. Theoretical  
555 and Applied Genetics **136**(7), 162 (2023). <https://doi.org/10.1007/s00122-023-04395-x> 555  
556 45. Sama, N., David, E., Rossetti, S., Antona, A., Franchetti, B., Pirri, F.: A new  
557 large dataset and a transfer learning methodology for plant phenotyping in vertical  
558 farms. In: Proceedings of the IEEE/CVF International Conference on Computer  
559 Vision. pp. 540–551 (2023) 559  
560 46. Sauvalle, B., de La Fortelle, A.: Autoencoder-based background reconstruc-  
561 tion and foreground segmentation with background noise estimation. CoRR  
562 [abs/2112.08001](https://arxiv.org/abs/2112.08001) (2021), <https://arxiv.org/abs/2112.08001> 562  
563 47. Scharr, H., Minervini, M., Fischbach, A., Tsaftaris, S.: Annotated image datasets  
564 of rosette plants. Technical Report No. FZJ-2014-03837 (07 2014) 564  
565 48. Scharr, H., Minervini, M., French, A.P., Klukas, C., Kramer, D.M., Liu, X., Luengo,  
566 I., Pape, J.M., Polder, G., Vukadinovic, D., et al.: Leaf segmentation in plant  
567 phenotyping: a collation study. Machine vision and applications **27**, 585–606 (2016)  
568 1, 3 568  
569 49. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale  
570 image recognition. Proceedings of the International Conference on Learning Rep-  
571 resentations (ICLR) (2015), <https://arxiv.org/abs/1409.1556> 13 570  
572 50. Singh, D., Jain, N., Jain, P., Kayal, P., Kumawat, S., Batra, N.: Plantdoc: A dataset  
573 for visual plant disease detection. In: Proceedings of the 7th ACM IKDD CoDS  
574 and 25th COMAD. p. 249–253. CoDS COMAD 2020, Association for Computing  
575 Machinery, New York, NY, USA (2020). [https://doi.org/10.1145/3371158.](https://doi.org/10.1145/3371158.3371196)  
576 [3371196](https://doi.org/10.1145/3371158.3371196), <https://doi.org/10.1145/3371158.3371196> 3, 4, 13 576  
577 51. Stavness, I., Giuffrida, V., Scharr, H.: Computer vision in plant phenotyping and  
578 agriculture. Frontiers in Artificial Intelligence **6**, 1187301 (2023) 1 578  
579 52. Szegedy, C., Liu, W., Jia, Y., et al.: Rethinking the inception architecture for  
580 computer vision. Proceedings of the IEEE Conference on Computer Vision and  
581 Pattern Recognition (CVPR) pp. 2818–2826 (2016). <https://doi.org/10.1109/CVPR.2016.308> 13 580  
582 53. Szegedy, C., Vanhoucke, V., Ioffe, S., et al.: Inception-v4, inception-resnet and the  
583 impact of residual connections on learning. Proceedings of the IEEE Conference on  
584 Computer Vision and Pattern Recognition (CVPR) pp. 4278–4286 (2017). <https://doi.org/10.1109/CVPR.2017.457> 13 584  
585 54. Triki, A., Bouaziz, B., Gaikwad, J., Mahdi, W.: Deep leaf: Mask r-cnn based leaf  
586 detection and segmentation from digitized herbarium specimen images. Pattern  
587 Recognition Letters **150**, 76–83 (2021) 1 586  
588 55. Vu, T., Jain, H., Bucher, M., Cord, M., Pérez, P.: ADVENT: adversarial en-  
589 tropy minimization for domain adaptation in semantic segmentation. CoRR  
590 [abs/1811.12833](https://arxiv.org/abs/1811.12833) (2018), [http://arxiv.org/abs/1811.12833](https://arxiv.org/abs/1811.12833) 3 590  
591 56. Walter, J., Edwards, J., Cai, J., McDonald, G., Miklavcic, S.J., Kuchel, H.: High-  
592 throughput field imaging and basic image analysis in a wheat breeding programme.  
593 Frontiers in Plant Science **10** (2019). <https://doi.org/10.3389/fpls.2019.00449>  
594 <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2019.00449> 1 594  
595 596 597 598

- 599 57. Wang, J., Cohen, M.F., et al.: Image and video matting: a survey. Foundations  
600 and Trends® in Computer Graphics and Vision **3**(2), 97–175 (2008) 5  
601 58. Wang, Y., Zhang, P., Tian, S.: Tomato leaf disease detection based on attention  
602 mechanism and multi-scale feature fusion. Frontiers in Plant Science **15**, 1382802  
603 (2024) 12  
604 59. Weickert, J.: A review of nonlinear diffusion filtering. In: International conference  
605 on scale-space theories in computer vision. pp. 1–28. Springer (1997) 7  
606 60. Weickert, J.: Coherence-enhancing diffusion filtering. International journal of com-  
607 puter vision **31**, 111–127 (1999) 7  
608 61. Weickert, J., et al.: Anisotropic diffusion in image processing, vol. 1. Teubner  
609 Stuttgart (1998) 7  
610 62. Weyler, J., Magistri, F., Marks, E., Chong, Y.L., Sodano, M., Roggiolani, G., Che-  
611 brolu, N., Stachniss, C., Behley, J.: Phenobench: A large dataset and benchmarks  
612 for semantic image interpretation in the agricultural domain. IEEE Transactions  
613 on Pattern Analysis and Machine Intelligence (2024) 3  
614 63. Weyler, J., Magistri, F., Seitz, P., Behley, J., Stachniss, C.: In-field phenotyping  
615 based on crop leaf and plant instance segmentation. In: 2022 IEEE/CVF Winter  
616 Conference on Applications of Computer Vision (WACV). pp. 2968–2977 (2022).  
617 <https://doi.org/10.1109/WACV51458.2022.00302> 3  
618 64. Xia, X., Kulis, B.: W-net: A deep model for fully unsupervised image segmentation.  
619 CoRR **abs/1711.08506** (2017), <http://arxiv.org/abs/1711.08506> 3  
620 65. Xie, S., Tu, Z.: Holistically-nested edge detection. In: Proceedings of the IEEE  
621 international conference on computer vision. pp. 1395–1403 (2015) 7, 9  
622 66. Zhang, J., Xie, T., Yang, C., Song, H., Jiang, Z., Zhou, G., Zhang, D., Feng, H.,  
623 Xie, J.: Segmenting purple rapeseed leaves in the field from uav rgb imagery using  
624 deep learning as an auxiliary means for nitrogen stress detection. Remote Sensing  
625 **12**(9), 1403 (2020) 1  
626 67. Zhang, S., Wang, H., Huang, W., You, Z.: Plant diseased leaf segmentation and  
627 recognition by fusion of superpixel, k-means and phog. Optik **157**, 866–872 (2018).  
628 <https://doi.org/https://doi.org/10.1016/j.ijleo.2017.11.190>, <https://www.sciencedirect.com/science/article/pii/S0030402617316108> 3  
629