

AutoCount: Unsupervised Segmentation and Counting of Organs in Field Images

Jordan R. Ubbens, Tewodros W. Ayalew, Steve Shirtliffe, Anique Josuttes,
Curtis Pozniak, and Ian Stavness

University of Saskatchewan, Saskatoon SK, S7N 5A8, Canada
{jordan.ubbens,tewodros.ayalew,steve.shirtliffe,
acj450,curtis.pozniak,ian.stavness}@usask.ca}

Abstract. Counting plant organs such as heads or tassels from outdoor imagery is a popular benchmark computer vision task in plant phenotyping, which has been previously investigated in the literature using state-of-the-art supervised deep learning techniques. However, the annotation of organs in field images is time-consuming and prone to errors. In this paper, we propose a fully unsupervised technique for counting dense objects such as plant organs. We use a convolutional network-based unsupervised segmentation method followed by two post-hoc optimization steps. The proposed technique is shown to provide competitive counting performance on a range of organ counting tasks in sorghum (*S. bicolor*) and wheat (*T. aestivum*) with no dataset-dependent tuning or modifications.

Keywords: Computer vision, organ counting, unsupervised segmentation.

1 Introduction

Object counting is a common real-world use case for deep learning and computer vision. In plant science and agriculture in particular, the counting of plant organs such as heads, tassels, flowers, pods, or fruiting bodies in field environments is relevant, as the number of organs correlates with agronomically important factors such as grain yield. Organ counting is a labor-intensive manual task which is practical to replace with automated methods. There is a substantial body of existing literature describing the counting of organs via supervised learning methods [11,9,10,2,14]. Among the most common supervised techniques are counting by density estimation or local counts regression, object detection, and counting by segmentation. The majority of these supervised methods were originally designed for much more difficult multi-class object detection and object counting problems, such as the challenging COCO benchmark [8]. In contrast, plant organs have low intra-class variability, but are dense and overlapping. Supervised methods also require data annotation in the form of either point annotations, bounding boxes, or segmentation masks. Collecting such annotations is difficult and time-consuming, given the typically high density of organs in

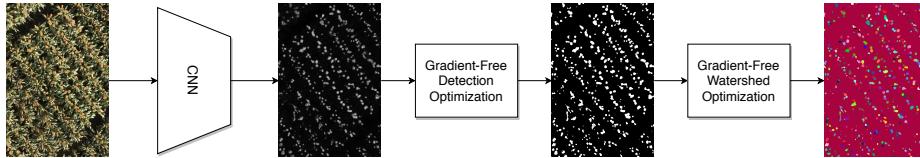


Fig. 1: The proposed process from image to instance segmentation.

field images. In addition, poor quality labels can contribute to non-convergence and overfitting as the label noise overpowers the signal in the annotations. Label quality has been shown to be degraded when annotations are provided by domain non-experts, who may be unfamiliar with the plant structures in question [3]. Principally, these supervised methods need to be retrained on annotated data for each unique dataset, because of the issue of domain shift [15,17].

In this paper, we propose a novel unsupervised counting technique for counting plant organs in aerial images of fields. We show that a combination of unsupervised segmentation using a CNN, a gradient-free optimization of threshold parameters, and a search over watershed segmentation parameters is able to perform well in these tasks. A complete diagram of the process is shown in Figure 1. We call this proposed object counting system *AutoCount*. We report unsupervised object counting results for two public datasets of sorghum heads (botanical term panicles), as well as a new dataset of wheat heads (botanical term spikes). These results represent new baselines for performing plant organ counting in a fully unsupervised setting, making organ counting easy to apply routinely on a wide variety of aerial image datasets.

1.1 Related Work

Organ counting is a common task in image-based plant phenotyping. Early organ counting efforts focused on leaf counting in *A. thaliana* rosettes grown in controlled, indoor conditions such as growth chambers and greenhouses [4,1,16]. Other work with indoor imaging data has localized wheat spikes and spikelets [12,11]. While performing organ counting with controlled imaging is relevant for many experiments, organ counting is also more relevant in outdoor field-based experiments and is frequently assessed using less-controlled imaging techniques such as with unmanned aerial vehicles (UAVs). Vision tasks in outdoor contexts are substantially different than those in indoor contexts, as factors such as lighting and plant motion cannot be controlled.

Previous work in organ counting in outdoor field environments has explored both image processing as well as deep learning methods. [5] used a decision tree-based segmentation with an image processing feature set, followed by the classification of regions using Bag-of-Visual-Words features for head detection in sorghum. Deep learning has provided some of the strongest performance for plant organ counting tasks in field imagery. Notably, [9] used local counts regression on a density map estimated by a CNN. The network showed strong performance for

counting the tassels of maize plants from field images. A follow-up work extended the technique to the counting of wheat heads [18]. Organ counting has also been explored using object detection methods, such as R-CNN [6], Faster R-CNN [10], as well as the state-of-the-art RetinaNet meta-architecture [2]. Others have demonstrated counting by segmentation. [14] proposed a superpixel segmentation followed by further segmentation using a fully convolutional network. A similar pipeline has been used for rice panicles [19].

2 Unsupervised Instance Segmentation

In contrast to existing organ counting methods, we propose a strictly unsupervised technique which does not require any annotated data. The technique is agnostic to any particular dataset, and does not require any hand-tuning to apply. An open source implementation is available¹. In the initial step, we make use of an existing unsupervised segmentation technique to train a segmentation network on the entire dataset [7]. This method utilizes a Simple Linear Iterative Clustering (SLIC) superpixel segmentation to first roughly segment the image. Next, a fully convolutional network consisting of three blocks of only convolution and batch normalization, is used to assign class labels to pixels. This network uses a stride of one on each convolutional layer to maintain the full size of the input image through the network. Pixel-level labels are assigned to superpixels based on the maximum class membership within each region, and the segmentation network is trained on this target using cross-entropy loss. Because of the trivial clustering solution where every region is assigned the same label, intra-axis normalization is used to encourage a larger number of unique class labels. This process is repeated until either a maximum number of iterations or a minimum number of unique class labels has been attained.

Next, the trained segmentation network is used to perform inference on each image in the dataset and the softmax function is applied to the raw activations of the final layer. The activations for the organ class in the output are extracted. The index of this label can either be determined by selecting the label which attains the lowest value of the objective function in Equation 2, or it can be indicated by the user. The result is a probability density map for the organ class for each image in the dataset.

In order to obtain a binary segmentation, a gradient-free optimization is performed on the activations using the portfolio-discrete One Plus One optimizer implemented in the Nevergrad library [13]. We find that this optimization procedure consistently provides more accurate segmentations of object regions than the class predictions from the segmentation CNN. Since the pixel values in this map correspond to probabilities, a log-likelihood loss function is optimized over individual activation maps $\mathbf{k}_i \in \{\mathbf{k}_0 \dots \mathbf{k}_N\}$ as

$$\operatorname{argmin}_{\theta} -\log \left(\frac{\sum m_{\theta}(\mathbf{k}_i)}{\sum \mathbf{k}_i} \right) - \log \left(\frac{\sum m_{\theta}(\mathbf{k}_i)}{|\{m_{\theta}(\mathbf{k}_i)\}|} \right) \quad (1)$$

¹ <https://github.com/p2irc/autocount>

where m_θ is a binary masking function which applies a Gaussian filter, a pixel threshold, and then a morphological dilation and erosion to the selected areas. This function returns its input masked by the selected pixels. $|m_\theta|$ indicates the number of selected pixels in that masking, and $\theta = (\tau, \lambda, \phi_{erosion}, \phi_{dilation})$ are the standard deviation of the Gaussian filter, the pixel threshold value, and the diameter of a circular kernel for morphological erosion and dilation, respectively. The first term of the loss function encourages the optimization to include as much of the total probability density as possible, while the second term encourages the average probability of selected pixels to be high.

Finally, objects smaller than a pixels are removed, and watershed segmentation is performed using the distance transform of the binary mask with local maxima no closer than b pixels apart used as watershed markers. These parameters are optimized over binary masks as

$$\operatorname{argmin}_{a,b} \frac{1}{N} \sum_{i=0}^N [W(\mathbf{s}_i, \hat{\mathbf{s}}_i) + W(\hat{\mathbf{s}}_i, \mathcal{N}_{\hat{\mathbf{s}}_i}) + W(\Gamma_{\hat{\mathbf{s}}_i}, \mathcal{N}_{\hat{\mathbf{s}}_i})]^2 \quad (2)$$

where \mathbf{s} and $\hat{\mathbf{s}}$ are the empirical distributions of object sizes before and following the procedure, W is the Wasserstein distance, $\mathcal{N}_{\hat{\mathbf{s}}} = \mathcal{N}(\mu(\hat{\mathbf{s}}), \sigma(\hat{\mathbf{s}})^2)$ is the normal distribution parameterized by $\hat{\mathbf{s}}$, and $\Gamma_{\hat{\mathbf{s}}} = \Gamma(\alpha(\hat{\mathbf{s}}), \beta(\hat{\mathbf{s}}))$ is the gamma distribution parameterized by $\hat{\mathbf{s}}$. This optimization is performed via an exhaustive search, because of the low dimensionality of the search space. In contrast to Equation 1 which is optimized on a per-sample basis, Equation 2 is optimized over the entire dataset. We find that this mitigates outlier samples which would otherwise find suboptimal solutions. Following optimization, the number of objects is reported as the number of object regions in the watershed segmentation.

Intuitively, for objects such as plant organs from top-down aerial images, we assume that the object sizes should be roughly normally distributed. The tension between the first term, which acts as a prior, and the second two terms, which shape the distribution by affecting its normality and skewness, finds a balance between over-segmentation and under-segmentation of object regions. Note that, if \mathbf{s} is already normally distributed, then the optimal solution to Equation 2 is to leave the distribution unchanged. Figure 2 shows the distributions of object sizes for the sorghum and wheat experiments before and after this step.

Although the training of the CNN in the initial step includes several tunable hyperparameters such as the learning rate and the number of convolutional blocks, we use the default settings provided by the authors' implementation for all experiments [7]. We adjust the number of output channels to 32 and the final number of labels (used as a stopping criteria) to eight, based on the low level of object diversity present in field datasets.

3 Results

We evaluate the counting performance of the proposed method using the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE), and the coefficient

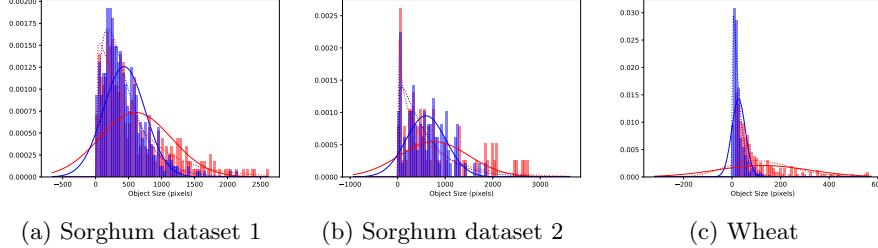


Fig. 2: Distributions of object sizes for a single example image before (red) and after (blue) the segmentation step. A normal distribution fit to the samples is shown as a solid line and a gamma distribution is shown as a dashed line.

of determination between the annotated and predicted counts (R^2). Table 1 shows a comparison against two previously reported results from the literature on two publicly available datasets of sorghum heads [5]. The resolution for these datasets is 1154 by 1731 pixels for dataset 1 and 1394 by 357 pixels for dataset 2. Although counting performance fell short of the previous standard set by the powerful RetinaNet object detection meta-architecture applied in [2], AutoCount shows that a high degree of performance can be attained without using labels. Examples from the public sorghum datasets are shown in Figure 3.

Table 1: Counting results for the sorghum and wheat head datasets.

		MAE	RMSE	R^2
Sorghum 1	Segmentation + Classification [5]	-	-	0.84
	RetinaNet [2]	-	-	0.82
	AutoCount (unsupervised)	29.92	36.39	0.79
Sorghum 2	Segmentation + Classification [5]	-	-	0.56
	RetinaNet [2]	-	-	0.76
	AutoCount (unsupervised)	4.95	6.22	0.48
Wheat	AutoCount (unsupervised)	297.22	344.89	0.17

The proposed method was also tested on a dataset of wheat field images (Table 1). The wheat head dataset consisted of 81 images from a *T. aestivum* breeding trial captured with a camera extended from an all terrain vehicle pointed downward at a height of approximately 6 m. The images were captured in RAW format with a 18 MP camera with an ultra-wide angle rectilinear lens. Wheat head centers were annotated by domain experts using a custom tool and the total number of boxes was used for the annotated count value. The images were

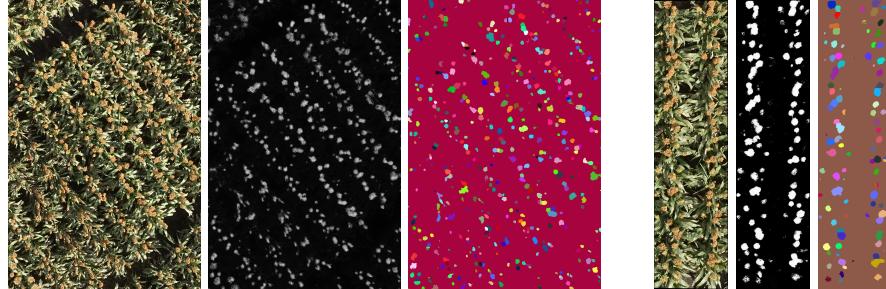


Fig. 3: Example images and intermediate outputs from the sorghum datasets. Left to right: original image, softmax activations, watershed output.

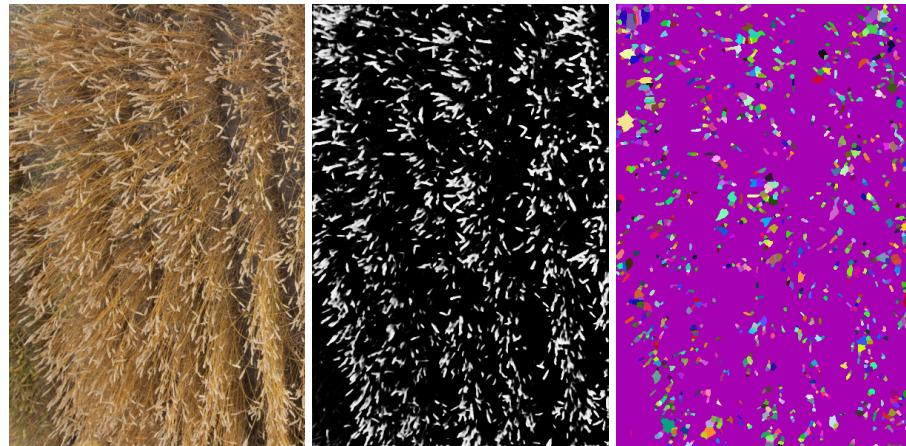


Fig. 4: Example image and intermediate outputs from the wheat dataset. Left to right: original image, softmax activations, watershed output.

resized to 400 by 600 pixels for expediency because of the extremely high resolution, although this resizing is not required and the original resolution can be used as was the case for the sorghum datasets. Compared to the sorghum dataset, the wheat images included substantially more visual variation among images, including different head densities, different levels of maturity, and both awned and unawned heads. Figure 4 shows an example from the wheat dataset. Despite this variation, AutoCount was applied to the wheat dataset with no modifications. Although the detection of heads is positive, the arrangement of the heads in dense clumps proved difficult to segment accurately using the watershed method, resulting in a lower R^2 value than in the sorghum experiments. This issue is discussed further in Section 4.

4 Discussion

AutoCount is best suited to datasets where organs are represented as contiguous regions, making aerial imagery a good candidate. In contrast, high-resolution proximal imaging could create a scenario where the organ region is over-segmented, for example, by segmenting a wheat head into individual spikelets. The counting performance of the method is expected to suffer under these conditions. The method also is subject to the weaknesses of watershed segmentation. One such drawback is the potential over-segmentation of objects which are not compact. For example, the performance of the method suffered significantly in the second dataset of sorghum due to a small subset of images containing a high density of ring-shaped panicles. An example of this mode of failure is shown in Figure 5. Another example of a failure case for the watershed method can be seen in the dense clumps of wheat heads shown in Figure 4. Many of these regions do not admit a shape which can be accurately segmented by the watershed method. Without the ability to accurately separate these large regions, the optimization of segmentation parameters ultimately settles on an over-segmentation of the wheat heads as can be seen in Figure 2c.

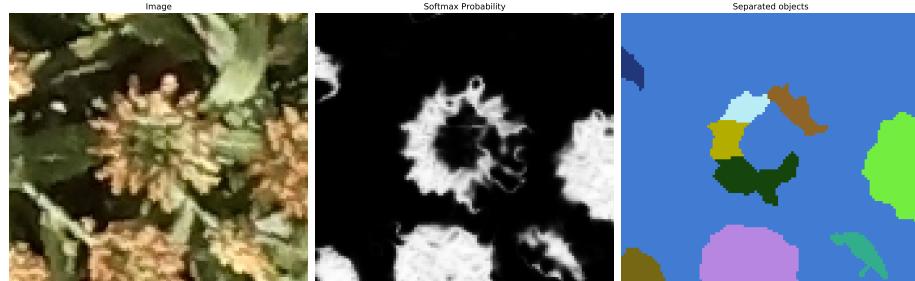


Fig. 5: A failure case from the second sorghum dataset. Ring-shaped organs are over-segmented by the watershed method.

Although we focus on the organ count in the present work as it is the most agronomically relevant trait, the method also provides an instance segmentation of organs by way of the final watershed segmentation following optimization. This mask could potentially be used for other applications such as measuring organ size. Future work could potentially address replacing the watershed method with a more robust segmentation technique, such as one that is able to act on the activation map directly as opposed to requiring a thresholded input. We expect that a more robust segmentation technique would help boost performance for wheat head images and for other challenging object instances which tend to become over-segmented with our current approach, and could be dropped into the overall system without the need to change the segmentation loss function.

5 Conclusion

Here we have introduced the first unsupervised object counting algorithm specifically tailored for counting plant organs in field environments, called AutoCount. Although it does not rely on labelled data, the counting performance of the method is comparable to that of previously reported supervised methods on two datasets of sorghum. Additional positive results are presented on a new dataset of wheat heads. Some limitations of the method and avenues for future work are identified. The results presented here represent the first fully unsupervised baselines for performing plant organ counting of arbitrary plant organs in diverse field imagery.

Acknowledgments This research was funded by a Canada First Research Excellence Fund grant from the Natural Sciences and Engineering Research Council of Canada.

References

1. Dobrescu, A., Giuffrida, M.V., Tsafaris, S.A.: Leveraging multiple datasets for deep leaf counting. In: Proceedings of the IEEE International Conference on Computer Vision Workshops. pp. 2072–2079 (2017)
2. Ghosal, S., Zheng, B., Chapman, S.C., Potgieter, A.B., Jordan, D.R., Wang, X., Singh, A.K., Singh, A., Hirafuji, M., Ninomiya, S., Ganapathysubramanian, B., Sarkar, S., Guo, W.: A Weakly Supervised Deep Learning Framework for Sorghum Head Detection and Counting. *Plant Phenomics* **2019**, 1–14 (2019). <https://doi.org/10.34133/2019/1525874>
3. Giuffrida, M.V., Chen, F., Scharr, H., Tsafaris, S.A.: Citizen crowds and experts: observer variability in image-based plant phenotyping. *Plant Methods* **14**(1), 12 (2018)
4. Giuffrida, M.V., Minervini, M., Tsafaris, S.A.: Learning to count leaves in rosette plants. In: Proceedings of the Computer Vision Problems in Plant Phenotyping (CVPNP) 2015. BMVA press (2016)
5. Guo, W., Zheng, B., Potgieter, A.B., Diot, J., Watanabe, K., Noshita, K., Jordan, D.R., Wang, X., Watson, J., Ninomiya, S., et al.: Aerial imagery analysis—quantifying appearance and number of sorghum heads for applications in breeding and agronomy. *Frontiers in Plant Science* **9**, 1544 (2018)
6. Hasan, M.M., Chopin, J.P., Laga, H., Miklavcic, S.J.: Detection and analysis of wheat spikes using Convolutional Neural Networks. *Plant Methods* **14**(1), 1–13 (2018). <https://doi.org/10.1186/s13007-018-0366-8>
7. Kanezaki, A.: Unsupervised image segmentation by backpropagation. In: 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP). pp. 1543–1547. IEEE (2018)
8. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: European conference on computer vision. pp. 740–755. Springer (2014)
9. Lu, H., Cao, Z., Xiao, Y., Zhuang, B., Shen, C.: TasselNet : Counting maize tassels in the wild via local counts regression network. *Plant Methods* pp. 1–14 (2017). <https://doi.org/10.1186/s13007-017-0224-0>

10. Madec, S., Jin, X., Lu, H., De Solan, B., Liu, S., Duyme, F., Heritier, E., Baret, F.: Ear density estimation from high resolution RGB imagery using deep learning technique. *Agricultural and Forest Meteorology* **264**(October 2018), 225–234 (2019). <https://doi.org/10.1016/j.agrformet.2018.10.013>
11. Pound, M.P., Atkinson, J.A., Wells, D.M., Pridmore, T.P., French, A.P.: Deep learning for multi-task plant phenotyping. In: Proceedings of the IEEE International Conference on Computer Vision Workshops. pp. 2055–2063 (2017)
12. Qiongyan, L., Cai, J., Berger, B., Okamoto, M., Miklavcic, S.J.: Detecting spikes of wheat plants using neural networks with laws texture energy. *Plant Methods* **13**(1), 83 (2017)
13. Rapin, J., Teytaud, O.: Nevergrad - A gradient-free optimization platform. <https://GitHub.com/FacebookResearch/Nevergrad> (2018)
14. Sadeghi-Tehran, P., Virlet, N., Ampe, E.M., Reynolds, P., Hawkesford, M.J.: Deep-Count : In-Field Automatic Quantification of Wheat Spikes Using Simple Linear Iterative Clustering and Deep Convolutional Neural Networks. *Frontiers in Plant Science* **10**(September), 1–16 (2019). <https://doi.org/10.3389/fpls.2019.01176>
15. Ubbens, J., Cieslak, M., Prusinkiewicz, P., Stavness, I.: The use of plant models in deep learning: an application to leaf counting in rosette plants. *Plant Methods* **14**(1), 6 (2018). <https://doi.org/10.1186/s13007-018-0273-z>
16. Ubbens, J.R., Stavness, I.: Deep plant phenomics: a deep learning platform for complex plant phenotyping tasks. *Frontiers in Plant Science* **8**, 1190 (2017)
17. Valerio Giuffrida, M., Dobrescu, A., Doerner, P., Tsafaris, S.A.: Leaf counting without annotations using adversarial unsupervised domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (2019)
18. Xiong, H., Cao, Z., Lu, H., Madec, S., Liu, L., Shen, C.: TasselNetv2 : in field counting of wheat spikes with context augmented local regression networks. *Plant Methods* (2019). <https://doi.org/10.1186/s13007-019-0537-2>
19. Xiong, X., Duan, L., Liu, L., Tu, H., Yang, P., Wu, D., Chen, G., Xiong, L., Yang, W., Liu, Q.: Panicle-SEG: a robust image segmentation method for rice panicles in the field based on deep learning and superpixel optimization. *Plant Methods* **13**(1), 104 (2017)