DATA DESCRIPTION AND FURTHER DETAILS

LEAF COUNTING CHALLENGE (LCC-2017)
COMPUTER VISION PROBLEMS IN PLANT PHENOTYPING (CVPPP) AT ICCV 2017

Venice, Italy, 28. October 2017, in conjunction with ICCV 2017 (http://iccv2017.thecvf.com/)

http://www.plant-phenotyping.org/CVPPP2017-challenge

1. Introduction

You are here because you are interested in submitting a contribution to the Leaf Counting Challenge component of the CVPPP workshop. If you have not done so please register with the challenge and read the instructions at:
https://www.plant-phenotyping.org/CVPPP2017-challenge

This document provides more information related to how the data were acquired, annotated, and how the testing and evaluation will occur.

This is the 2nd leaf counting challenge and it aims for finding solutions that can learn to count data directly via learning mechanisms. Authors could submit algorithms that aim to count via detection but note that we are not providing training data of that sort.

Why this challenge and how is it different from LSC?
From a phenotyping perspective, the number of leaves is directly related to yield potential, drought tolerance, and flowering time. From a vision perspective, it can also be used to constrain leaf detection or leaf segmentation algorithms. Usually, user interaction is required and leaf count comes as a by-product of leaf segmentation. Learning-based counting techniques could help here and for that reason we would like to collect approaches that address this area. It is different from LSC due to how the supervision is given. In LSC per leaf segmentations are provided (a much stronger and informative annotation, albeit more laborious) whereas here we provide per-image count and dot-annotations. Still note that the labeled data offer many ways of building algorithms: directly as regression problems [eg. see Minervini et al 2017, http://onlinelibrary.wiley.com/doi/10.1111/tpj.13472/abstract ; Giuffrida et al 2015 https://dx.doi.org/10.5244/C.29.CVPPP.1], or learning densities via leaf center annotations. In addition, we provide foreground masks should the participants like to use them.

2. Data Description

2.1. Overall

These data have been collected from several sites from growth chamber experiments. The dataset contains images of tobacco and Arabidopsis plants in separate folders.

Tobacco images (**folder A3**) were collected using a camera, which contained in its field of view a single plant.

Arabidopsis images were collected using a camera with a larger field of view encompassing many plants, which were later cropped. The images released are either from mutants or wild types and have been taken in a span of several days and are from two different experimental setups folders **A1** and **A2** and were the field of view is different. Furthermore, due to the broader field of view some plants can be slighter out of focus than others. Finally, note that while in most images the background is simple and static, in some cases the growth of moss or the presence of water in the growing tray complicate the scene. These images are introduced to demonstrate the complexity of the problem in the context of foreground/background segmentation. All images were hand labeled to obtain ground truth masks for each leaf in the scene. Examples of raw and labeled images are in Figure 1. Based on user annotations we found leaf centers and leaf counts.

We also include data derived from a public dataset (original data kindly shared by Dr Hannah Dee from Aberystwyth), here providing leaf counts with cropped images.

For further information on the data sources please refer to:

1. M. Minervini, A. Fischbach, H.Scharr, and S.A. Tsaftaris. Finely-grained annotated datasets for image-based plant phenotyping. Pattern Recognition Letters, pages 1-10, 2015, doi:10.1016/j.patrec.2015.10.013 [PDF] [BibTex]
2. Bell, Jonathan, & Dee, Hannah M. (2016). Aberystwyth Leaf Evaluation Dataset [Data set]. Zenodo. http://doi.org/10.5281/zenodo.168158

File types and naming conventions: Plant images are encoded as PNG files and their size may vary. Plants appear centered in the cropped image (whenever possible). Segmentation masks are image files encoded in PNG where each segmented leaf is identified with a unique (per image) integer value, starting from 1, where 0 is background. A color index palette is included within the file for visualization reasons. The filenames have the form:

plantXXX_rgb.png	➔	the raw color image in RGB;
plantXXX_centers.png	➔	a labeled image as indexed PNG file, with a single pixel denoting a leaf center;
plantXXX_fg.png	➔	the foreground (plant segmentation) as binary PNG file;

where XXX is a 3 or 4 digit integer number. Note that plants are not numbered continuously.

In addition, we provide:

AY.csv ➔ a CSV file listing image names and number of leaves is also provided, for convenience of approaches that solve directly the regression problem.
where Y is 1, 2,3 or 4 to denote A1, A2, A3 or A4.

2.2. Training set

We provide 27 images of tobacco and 783 Arabidopsis images to the registered users, and 4 CSV files containing ground truth leaf counts for all images.

2.3. Testing set

Registered authors will receive **June 1st** the testing set(s) for which we provide plant images and their foreground segmentation**.** However, we will not share ground truth leaf counts or centroids.
We intend to share two different versions of the testing set:

1. [SPLIT] images are split according to the origin i.e. following the A1,…, A4 nomenclature.
2. [WILD] images are included in one folder (A5) only and may vary in size. This tries to emulate a leaf counting in the wild scenario where data from different sources are

pooled in the testing phase. If you want to perform well in this testing set we advise that you aim to pool data from A1 to A4 together.

Additional information will be provided in due time.

Please note that **IT IS STRICTLY FORBIDDEN** to attempt to use the testing set in any other manner, e.g., to label testing data for improved training, to check algorithmic performance visually on the testing data, etc. The organizers reserve the right to release a new testing set prior to the challenge for verifying the reported average performance of participants.

3. Evaluation Function

We will use the evaluation function `LCC_evaluation.m` (in MATLAB) we share with you in the Matlab archive for comparing segmentation outcomes between ground truth and algorithm results. The function uses count differences to evaluate counting results. It returns measures described in Table 1, shown below.

Table 1. Reported values of the evaluation functions

Metric		**Purpose**
CountDiff:	Returns the difference in leaf count, as number of leaves in the algorithms results minus the ground truth. On a set we measure average count and standard deviation.	To estimate how good the algorithm is in identifying the correct number of leaves present (to estimate overall bias)
AbsCountDiff:	Returns the absolute difference in leaf count, as number of leaves in the algorithms results minus ground truth. On a set we measure average count and standard deviation.	To estimate how good the algorithm is in identifying the correct number of leaves present (to estimate absolute errors)
CountAgreement:	Binary variable of correct counts, i.e. is 1 if no error was made (*CountDiff=0)*, and 0 otherwise. On a set we measure*PercentAgreement* i.e. total correct agreement obtained by summing the ones in *CountAgreement*, expressed in percentage.	To estimate in how many cases no count error was made.
MSE:	the average squared error of leaf count. Naturally we evaluate this for a set only.	This is used traditionally in regression problems and is adopted here for completion.

4. Evaluation Phase

Below we outline the evaluation process and expected results to be reported by contributors in their submitted papers. Note that these might be updated later on as we obtain feedback from the participants.

No later than **June 20th 2017**, authors will submit their results as a single ZIP archive named as: LastNameRegisteredAuthor_FirstInitial_results.zip
(e.g., Tsaftaris_S_results.zip) via email to a predefined email address. Inside the ZIP file, the following files are requested:

> *AY.csv*➔ a CSV file where per each row image name a leaf count is provided, one for each A1, A2, A3 and A4. An estimated count value must be provided for each of the testing images (i.e. no missing values).

Additional information will be shared to the registered author(s) in due time.

Within 48 hours the registered author will receive the results of the evaluation on a per testing image base.

Submission: Authors in their paper should report averages (and standard deviation) of the values obtained of the evaluation functions (Table 1). They should report results averaged across all three experiments but also individually on the training and testing set. They are also encouraged to report cases where their proposed algorithm did not perform as expected. They are required to mention explicitly:

a) if they use leaf centroid information;
b) if the use the foreground plant mask;
c) if they are dealing with each of A1 to A4 independently, i.e. a customized approach for each dataset;
d) if their method obtains counts via segmentation/detection or via learning;
e) their performance in the testing set [SPLIT and/or WILD]

5. Acknowledgements

Financial support of the EU Project PHIDIAS(http://prian.lab.imtlucca.it/PROJECTS/PHIDIAS/phidias.html), the DPPN(http://www.dppn.de/dppn/DE/Home/home_node.html), and EPSRC (http://gtr.rcuk.ac.uk/projects?ref=studentship-1785349) is acknowledged.