DATA DESCRIPTION AND FURTHER DETAILS
LEAF SEGMENTATION CHALLENGE (LSC-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 Segmentation 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 3rd LSC after the successful LSC 2014 and 2015. Examples of methods stemming from these challenges or using the data are http://link.springer.com/article/10.1007/s00138-015-0737-3 , https://arxiv.org/abs/1605.09410 , https://arxiv.org/abs/1511.08250).

The major difference of this challenge is the expansion of the data and that we focus on leaf segmentation accuracy and as such ground truth foreground segmentation masks are provided for training and testing.

2. Data Description

2.1. Overall

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.

We also include data derived from a public dataset (original data kindly shared by Dr Hannah Dee from Aberystwyth).

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. 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_label.png ➔ the labeled image as indexed PNG file
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.

2.2. Training set

We provide 27 images of tobacco and 783 Arabidopsis images and label images to the registered users.

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 segmentations.

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.

Table 1. Reported values of the evaluation functions

Function name		Purpose
SymmetricBestDice:	Symmetric Average DICE among all objects (leaves)	to estimate average leaf segmentation accuracy
FGBGDice:	DICE on the foreground mask (i.e. the whole plant assuming the union of all labels different than background)	to estimate how good the algorithm identifies plant from background. Note this metric will not be used for evaluation.We are not going to care for foreground segmentation quality for LSC 2017, since ground truth masks are made available.
AbsDiffFGLabels:	Returns the absolute difference in object count, as number of leaves of the algorithm's results minus the ground truth	to estimate how good the algorithm is in identifying the correct number of leaves present
DiffFGLabels:	Returns the difference in object count, as number of leaves of the algorithm's results minus the ground truth	to estimate how good the algorithm is in identifying the correct number of leaves present

3. Evaluation Function

We will use the evaluation function `LSC_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 the Dice function to evaluate segmentation results. It returns measures described in Table 1, shown previously.

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 community.

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 folder structure of the testing dataset should be maintained but should include only the results of your algorithms, the filenames should be:

> *plantXXX_label.png* ➜ the labeled image as indexed PNG file

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

Within 48 hours the registered authors 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 whether they used the SPLIT or WILD setting or both for testing. They should report results averaged across all three experiments but also individually. They are also encouraged to report cases where their proposed algorithm did not perform as expected.

5. Acknowledgements