Pheno-Deep Counter: the universal and versatile deep learning architecture for leaf count

Mario Valerio Giuffrida[1,2]
V.Giuffrida@ed.ac.uk
Peter Doerner[3]
Peter.Doerner@ed.ac.uk
Sotirios A. Tsaftaris[1,4]
S.Tsaftaris@ed.ac.uk

[1] Institute of Digital Communications,
School of Engineering,
The University of Edinburgh,
Edinburgh, EH9 3FG, UK

[2] IMT School for Advanced Studies,
P.zza San Francesco, 19
55100, Lucca (LU), Italy

[3] School of Biological Sciences,
The University of Edinburgh,
Edinburgh, EH9 3BF, UK

[4] The Alan Turing Institute,
London, NW1 2DB, UK

Abstract

The direct and manual observation of plants morphological traits is a tedious, time-consuming, and error-prone task. For this reason, there is an increasing interest in image-based plant phenotyping. However, the lack of robust and reliable software able to analyse plant images is causing a bottleneck [6]. The automatic extraction of plant traits is not easy, due to significant variability in the images, challenging the algorithms.

In this work, we are focused on the automatic prediction of the leaf count from top-view images of rosette-shaped plants. From a biological perspective, leaf count is an important trait to extract. In fact, it is directly related to the development stage of plants, can give an indication of the flowering time, and also allows the assessment of plant's health [4]. In this context, several nuisance factors have to be taken into account: leaves vary in scale, they occlude each other due to nastic movements, acquisition setup (e.g., cameras or illumination) is different.

In the past, leaf count was typically obtained by the per-leaf segmentation of plants. This kind of approaches finds a suitable delineation of each leaf, providing the leaf number by counting the number of segmented leaves. Overall, there approaches adopt heuristics, based on certain assumptions, that may fail when new variability is encountered on the data. For these reasons, machine learning algorithms are the new alternative in the plant community. Instead of developing ad-hoc software requiring specific conditions to operate, machine learning algorithms can be trained to solve specific tasks, by providing annotated data. Neural networks have shown superior performance to solve specific tasks (e.g., classification, segmentation, regression), compared with other machine learning approaches. Neural networks are able to extract high discriminative features from images, especially with deep architecture. However, such finely-grained annotated plant images

Table 1: Testing performance of *PhenoDC* on the multi-modal dataset [2]. We report results when the network is trained using only a single input and when also using all inputs.

Training on	DiC	|DiC|	MSE	%
RGB Only	0.02 (0.75)	0.48 (0.57)	0.56	55.7
FMP Only	-0.06 (0.72)	0.45 (0.56)	0.52	58.7
NIR Only	0.13 (0.61)	0.33 (0.53)	0.39	69.6
All	0.11 (0.40)	0.13 (0.39)	0.17	88.5

are hard to obtain and require a certain level of expertise of the annotator. Therefore, if one is interested in the total leaf count, a different approach can be considered.

Rather than segmenting each leaf individually (which requires specific annotations to successfully train a network), a regression model can predict the number of leaves of a given input image. In this case, the algorithm requires only a training set with images and total count, which is a rather simpler ground-truth to obtain. In [4], they proposed a leaf count algorithm, extracting a holistic plant descriptor learning a visual dictionary. Then, each plant of the dataset is encoded and an SVR is trained to infer the count of leaves. Deep neural networks can also be trained to output the total leaf count. In [1], they proposed an approach using two networks: the first one segments the plant from the background (per-plant segmentation), then the second network performs the leaf count from the segmented image. In [3], they proposed a deep neural network based on ResNet [5], training it with datasets of different sources.

Here, we present Pheno-Deep Counter, a single deep network that can predict leaf count in 2D plant images of different species with rosette-shaped appearance (c.f. Figure 1). We demonstrate that our architecture can count leaves from multi-modal 2D images, using RGB, fluorescence, or near-infrared images. Our network design is flexible, allowing for inputs to be added or removed to accommodate new modalities. In fact, we use a modality branch to extract meaningful features from each source and features are combined together in a common latent space with a max-pooling operation. Pheno-Deep Counter is able to produce accurate predictions in many plant species and, once trained, can count leaves in a few seconds

We trained our model on the multi-modal imagery [2]. Specifically, we used 576 annotated images from three modalities, generating four random splits for training (50%), validation (25%), and testing (25%) purposes. In Table 1, we report the experimental test results. Firstly, we trained our network using only one modality, where the best result was obtained with the near-infrared modality ($MSE = 0.39$). Training the network with three modalities simultaneously, it can be observed that the performance is considerably improved, reducing the MSE by $\sim 50\%$. This demonstrates that multi-modal data helps the learning of better features to obtain more accurate leaf counting predictions.

Our network is able to learn multi-modal features, processing each of the image modality. The features extracted from each modality are fused together, which are highly discriminative for the counting task. Since we demonstrated that multiple modalities improve performance, future work should involve the inference of multiple tasks, such as age or plant leaf area.

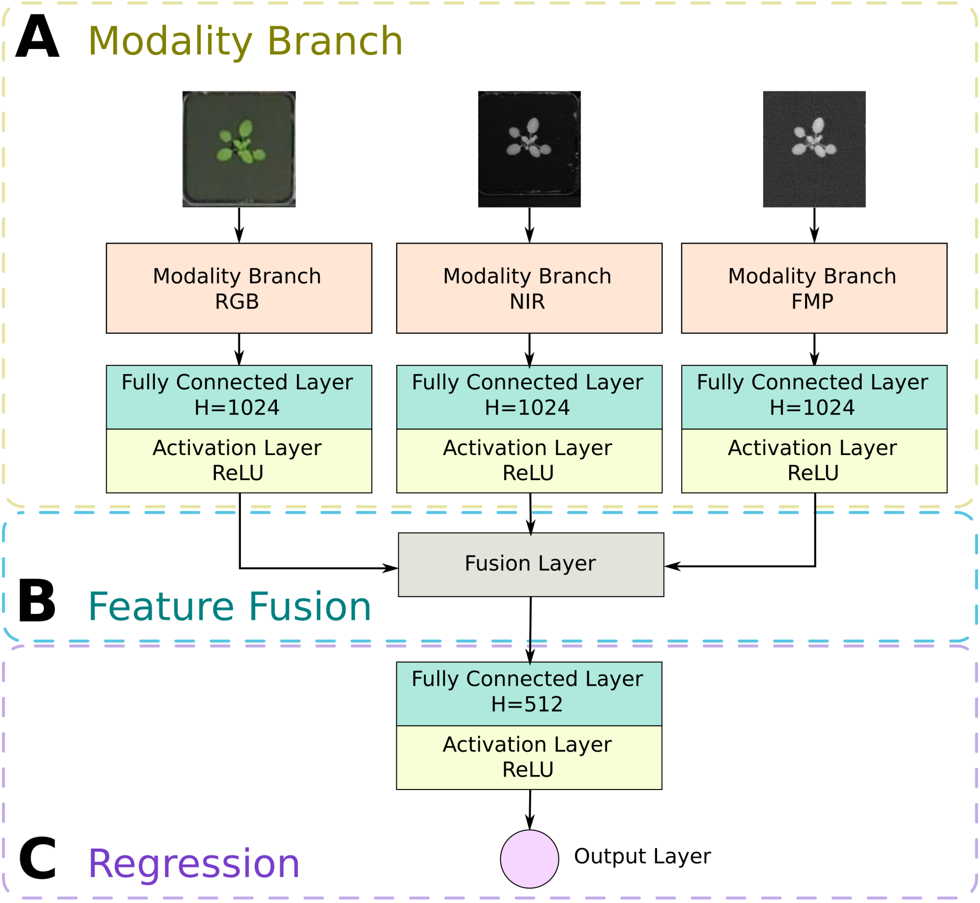

Figure 1: Schematic of the proposed deep architecture. (A) a modality branch, consisting of ResNet50 [5], extracts modality-dependent plant features as a feature vector of 1,024 neurons. (B) The fusion part combines those features to retain the most useful information from each modality. (C) The regression part, relates fused information with leaf count as a non-linear regression.

References

[1] Shubhra Aich and Ian Stavness. Leaf Counting with Deep Convolutional and Deconvolutional Networks. In *IEEE International Conference on Computer Vision Workshops Workshops*, pages 2080–2089. IEEE Computer Society, 2017. ISBN 978-1-5386-1034-3. doi: 10.1109/ICCVW.2017.244.

[2] Jeffrey A. Cruz, Xi Yin, Xiaoming Liu, Saif M. Imran, Daniel D. Morris, David M. Kramer, and Jin Chen. Multi-modality imagery database for plant phenotyping. *Machine Vision and Applications*, 27(5):735–749, jul 2016. ISSN 0932-8092. doi: 10.1007/s00138-015-0734-6.

[3] Andrei Dobrescu, Livia C T Scorza, Sotirios A Tsaftaris, and Alistair J McCormick. A "Do-It-Yourself" phenotyping system: measuring growth and morphology throughout the diel cycle in rosette shaped plants. *Plant Methods*, 13(1):95, nov 2017. ISSN 1746-4811.

[4] Mario Valerio Giuffrida, Massimo Minervini, and Sotirios Tsaftaris. Learning to Count Leaves in Rosette Plants. In *CVPPP workshop - BMVC*, page 13. British Machine Vision Association, 2015. ISBN 1-901725-55-3.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity Mappings in Deep Residual Networks. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 630–645, Cham, 2016. Springer International Publishing. ISBN 978-3-319-46493-0.

[6] Massimo Minervini, Hanno Scharr, and Sotirios A. Tsaftaris. Image analysis: The new bottleneck in plant phenotyping [applications corner]. *IEEE Signal Processing Magazine*, 32(4):126–131, 2015. ISSN 10535888.