Supplementary File S1:

Methodology of stomata detection and counting through machine learning.

Author: Dr. Soltani

Method

Our method is based on transfer learning for deep neural networks. That is, we have utilised a pre-trained deep model for the different datasets and adapt it for our stomata samples. Here, we briefly review the transfer learning technique and provide the details of the used model in our experiment.

Transfer learning

Transfer learning is a machine learning technique in which a learner learns a new task using the experiences of another task but related to the new task. It is also a popular approach in deep learning as it helps to significantly reduce the training time and complexity of designing a deep model resulting in a well-trained model for the target task with good generalisation performance. Transfer learning is used extensively in computer vision and natural language processing. For instance, if the target task is the classification of the CIFAR-10 images, it is better to use a pre-trained model on a very large dataset such as ImageNet as a starting point and adapt this pre-trained model for the new task (e.g., classifying of CIFAR-10 images) by only training a small part of the model. In this way, not only we can use the features learned by a related task, but also, we do not need to design and start training parameters of a model from scratch. This is indeed one of the important features of modern deep learning frameworks such as TensorFlow and PyTorch which provide a rich class of pre-trained models as a built-in API. This helps practitioners and researchers to quickly train an existing model for their own setup to test the performance.

Object Detection Model

Based on the transfer learning approach, we utilise a pre-trained object detection model trained on the standard COCO datasets (Lin *et al*., 2014). This dataset is a collection of more than 330k images with 80 object categories for large-scale object detection, segmentation, and captioning tasks. Since our goal is detecting and classifying stomata, we use the Faster R-CNN model (Ren *et al*., 2015) as one of the state-of-the-art methods based on deep neural networks. In

particular, we downloaded a pre-trained Faster R-CNN model available in Tensorflow with the Inception-V2 architecture (Szegedy *et* al., 2016) as the base model. Inception-V2 is a variation of Inception-V1 also referred to as GoogLeNet was the state-of-the-art architecture at ImageNet competition in ILSRVRC 2014. After loading the pre-trained Faster R-CNN, the last few layers of classification layers are changed to meet the aim of stomata classification and detection. In the next step, the Faster R-CNN with stomata images are trained with different hyper parameters such as learning rate and number of epochs to find out the best parameters to reduce execution time and errors.

References:

Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL. 2014. Microsoft coco: Common objects in context. *European conference on computer vision* **8693**: 740-755.

Ren S, He K, Girshick R, Sun J. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*: 91-99.

Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. 2016. Rethinking the inception architecture for computer vision. *Proceedings of the IEEE conference on computer vision and pattern recognition*: 2818-2826.