From leaf to label: a robust automated workflow for stomata detection

Authors:

Sofie MEEUS*
Meise Botanic Garden, Nieuwelaan 38, 1860 Meise, Belgium

Jan VAN DEN BULCKE
Department of Environment,
Ghent University, Coupure Links 653, 9000 Gent, Belgium

Francis WYFFELS
Department of Electronics and information systems, IDLab-AIRO,
Ghent University–imec, Technologiepark-Zwijnaarde 126, 9052 Zwijnaarde, Belgium

*corresponding author

Submitted: Monday 16th December, 2019

Abstract

1. Plant leaf stomata are the gatekeepers of the atmosphere-plant interface and are essential building blocks of land surface models as they control transpiration and photosynthesis. Although more stomatal trait data is needed to significantly reduce the error in these model predictions, recording these traits is time-consuming, and no standardized protocol is currently available. Some attempts were made to automate stomatal detection from photomicrographs, however, these approaches have the disadvantage of using classic image processing or targeting a narrow taxonomic entity which makes these technologies less robust and generalizable to other plant species. We propose an easy-to-use and adaptable workflow from leaf to label. A methodology for automatic stomata detection was developed using deep neural networks according to the state-of-the-art and its applicability demonstrated across the phylogeny of the angiosperms.

2. We used a patch-based approach for training/tuning three different deep learning architectures. For training, we used 431 micrographs taken from leaf prints made according to the nail polish method from herbarium specimens of 19 species. The best performing architecture was tested on 595 images of 16 additional species spread across the angiosperm phylogeny.

3. The nail polish method was successfully applied in 78% of the species sampled here. The VGG19 architecture slightly outperformed the basic shallow and deep architectures, with a confidence threshold equal to 0.7 resulting in an optimal trade-off between precision and recall. Applying this threshold, the VGG19 architecture obtained an average F-score of 0.87, 0.89, and 0.67 on the training, validation, and unseen test set, respectively. The average accuracy was very high (94%) for computed stomatal counts on unseen images of species used for

23 training.

24 4. The leaf-to-label pipeline is an easy-to-use workflow for researchers of different areas of

25 expertise interested in detecting stomata more efficiently. The described methodology was

26 based on multiple species and well-established methods so that it can serve as a reference for

27 future work.

Keywords

29 Deep neural networks, deep learning, detection, herbarium, optical microscope images, plants,

30 stomata, stomatal density, VGG19

31 ————————————————————————————————————-

1   Introduction

The study of ecosystem functioning requires a thorough understanding of the physiological processes of organisms occurring at the individual level. Organisms can be defined in terms of their functional traits, which are the phenotypic characteristics that are related to the fitness and performance of an organism. The spatial distribution of these functional traits in combination with environmental conditions, constitute the global diversity in ecosystem functioning and are therefore essential building blocks of land surface models (LSM). LSM are essential for estimating transpiration and photosynthesis from vegetated surfaces (Jefferson et al., 2017), the dominant component of global land evapotranspiration, and are a key component in models for operational predictions of the near-climate (Kushnir et al., 2019; Bertolino et al., 2019). Transpiration in an ecosystem, in essence, occurs at the individual leaf surface where stomata function as 'gates' between deep-soil water reservoirs and the atmosphere. Leaf stomata are microscopic pores surrounded by two guard cells ranging from approximately 10-100 μm in length. They control the balance between water loss and CO_2 uptake by the leaves and therefore have an important effect on the global carbon and hydrologic cycle (Berry et al., 2010; Steinthorsdottir et al., 2012; Wang et al., 2015). Moreover, as stomatal traits show a clear response to environmental parameters such as climate (e.g. Liu et al., 2018a) and atmospheric carbon dioxide concentrations (e.g. Woodward, 1987; Tanaka et al., 2013), they are key proxies of environmental change (Hetherington and Woodward, 2003). Stomatal conductance (gs), defined as the uptake rate of carbon dioxide or water vapor loss through the stomata of a leaf, is an elemental parameter in the LSM linking plant water use and carbon uptake (Kala et al., 2016) and is constrained by and derived from the size and density of the leaf stomata (Drake et al., 2013). It is well known that (maximum and minimum) stomatal conductance, as well as stomatal size, density, and rate of response, vary widely across plant species. Recent efforts have mapped stomatal

behavior globally (a.o. Lin et al., 2015), yet more detail is needed as including more interspecific trait variation in climate models could significantly reduce the error in model predictions (Wolz et al., 2017; Butler et al., 2017). To be useful in global-scale mapping, functional traits should be relatively easy and inexpensive to measure in a large number of taxa using a standardized protocol (Cornelissen et al., 2003; Perez-Harguindeguy et al., 2013; Moretti et al., 2017; Dawson et al., 2019). Recording stomatal traits is widely considered to be labor-intensive and time-consuming, and to this day, mostly performed manually (e.g., counting stomata through the microscope) and, therefore not replicable. The aspects of the methodologies currently used that add to the cost and intensity of the labor are the i) preparation of the leaves to be viewed with a microscope, ii) the number of replicates to account for the intra-individual variation in stomatal traits and iii) the measurements, either counts or size measurements, themselves. Few methods to automate the detection of and measurement on stomata have been reported in the literature, and in most cases, they consist of conventional image processing using algorithms that have to be tweaked to the specific task at hand. Scarlett et al. (2016) for instance, apply Maximum Stable External Regions to detect potential ellipses of stomata on microscope images of vine leaves while da Silva Oliveira et al. (2014) use Gaussian filtering and a series of morphological operations to detect stomata on optical microscope imagery of five different plant species. Duarte et al. (2017) use wavelet spot detection in tandem with standard image processing tools to segment stomata on one plant species and Higaki et al. (2014) combine a genetic algorithm and self–organizing maps, coined Clustering-Aided Rapid Training Agent, for the detection of stomata on fluorescently-labeled cell contour images of the leaf epidermis of *Arabidopsis* leaves. A series of other papers relies on classifiers for detecting of stomata. Vialet-Chabrand and Brendel (2014) report on the use of a cascade classifier for rapid assessment of the density and distribution of stomata on the leaves of two oak species. By training a Haar feature-based classifier with exemplary stomata, they can be detected with high accuracy on

SEM microphotographs. Jayakody et al. (2017) use a cascade object detection learning algorithm to correctly identify multiple stomata on rather large microscopic images of grapevine leaves, but also apply a combination of image processing techniques to estimate the pore dimensions of the stomata that were detected with the cascade object detector. Typically, the applied classic image processing techniques are based on handcrafted features for the detection and segmentation of the desired stomata. While these techniques perform well on one specific plant species, they do not generalize to other species.

An answer to the limitations of classical image processing techniques came from the field of neural networks with the introduction of deep learning. In a significant breakthrough, Krizhevsky et al. (2012) showed that deep learning was capable of achieving record-breaking results for object recognition. Deep learning allows computational models that are composed of multiple processing layers to learn representations from raw data with multiple levels of abstraction (LeCun et al., 2015; Najafabadi et al., 2015a). Since then, deep learning was quickly adopted by the vision community, which led to state-of-the-art results for the prediction of galaxy pictures (Dieleman et al., 2015), face recognition (Parkhi et al., 2015) or the detection of anatomical structures (Shen et al., 2017; Hoo-Chang et al., 2016). Its application is now being explored in different fields of biology including plant phenotyping (e.g. Jackson et al., 2017) and taxonomy (e.g. Wäldchen and Mäder, 2018), and very recent work has used deep learning for the detection of stomata (Fetter et al., 2018; Aono et al., 2019). LeCun et al. (2015) state that all of these successes in deep learning can be explained by the increase in computing power via GPUs, the ease with which data can be collected and various improvements for neural network techniques. Moreover, with the advent of deep learning toolboxes such as Keras (Chollet et al., 2015), deep learning also became accessible for non-computer scientists. Although deep learning can outperform other machine learning algorithms, training data is needed. Despite their important function, no standardized methodology has yet been described to measure

103   stomatal traits such as stomatal density and size. The handbook of protocols for the measurement

104   of plant functional traits (Cornelissen et al., 2003), highlights the importance of stomata as hard

105   functional traits, however, does not include any advice standardized way on how to prepare, image

106   and count them, while there is a clear need in the framework of global efforts on the one hand (Lin

107   et al., 2015), and to feed our deep learning networks on the other hand.

108   Finally, the recent paper by (Christin et al., 2019) highlights the importance of guidelines and

109   recommendations to help ecologists get started with deep learning. Although deep learning has

110   proven its potential in a lot of disciplines, developing a deep learning solution is not yet a trivial

111   task. They strongly advocate a stronger interaction between computer scientists and ecologists.

112   Here, we describe the development of deep learning models to perform stomatal detection/counts

113   automatically. To this end, we developed a pipeline of actions from leaf preparation to microscope

114   imaging that is easy, inexpensive, and acquires enough image quality to train and use the DL network.

115   The objective of this paper is, therefore twofold. (1) We provide a methodological protocol aimed

116   at standardising sample preparation as well as imaging of stomata. The rationale is to facilitate

117   comparability and usability across studies for revealing patterns and mechanisms by increasing the

118   reliability and predictive power of stomatal counts. More specifically, we outline an accessible

119   methodology to obtain stomatal counts "from leaf to label" that can be applied beyond a lab setting

120   and is also suitable for educational purposes. (2) We present a detailed and replicable methodology

121   for automatic stomata detection with deep neural networks and show its applicability of deep learning

122   across the phylogeny of the angiosperms. Our aim is to motivate researchers from the ecology and

123   evolution community to consider deep learning techniques for the automation of their workflows.

2 Materials and methods

2.1 Dataset generation

2.1.1 Specimens and species

We used mounted specimens from the African herbarium collection of Meise Botanic Garden which contains approximately 500.000 herbarium specimens from Burundi, Rwanda and Democratic Republic of the Congo, representing more than 80% of the existing collections from these countries (Stoffelen P., pers. comm., 2019). Five fully developed leaves per specimen were carefully detached and remounted afterwards. The species for algorithm training were selected in the context of studying the effects of global change on the central African forest vegetation. The specimens used here mainly came from common tropical (timber) tree species such as *Cola griseiflora*, *Mammea africana*, and *Erythrophleum suaveolens* which are well-represented in the collection and were recurrently collected throughout the last century (1902-2013) (for complete species list, see Suppl. Table). These herbarium specimens were collected at the Yangambi Biosphere Reserve, situated within the Congo River Basin west of the City Kisangani in the Democratic Republic of the Congo.

2.1.2 Leaf prints

Epidermal leaf impressions were made from the abaxial side of the leaves in the middle of the leaf, between the midvein and edge. This region of the leaf has been shown to contain guard cell lengths and stomatal densities comparable to the means of the entire leaf (see Beaulieu et al., 2008, and references therein). Transparent nail polish (Bourjois Cristal ball) was used to make the impressions which, once dried, were mounted pointing upward with double-sided tape (Scotch[TM]) on a microscope slide.

7

2.1.3   Image acquisition

Three photomicrographs of 1600x1200 pixels were taken per leaf print (dimensions = 344x258 μm; area view field = 0.09 mm²) using a digital microscope (VH-5000 Ver 1.5.1.1, KEYENCE CORPORATION, Osaka, Japan) with full coaxial lightning and default factory settings for shutter speed at ×1000 lens magnification (VH-Z250R). A single photomicrograph was created by stacking of several digital images taken at different focal planes to increase the depth of the resulting image. All stomata that fell entirely within the view field were counted and converted to stomata per square millimeter to obtain stomatal density.

2.2   Model development

2.2.1   Deep learning approach

A basic deep learning architecture is depicted in Fig. 1C. It consists, from left to right, of an input layer, multiple stacked convolutional and pooling layers, a fully connected feedforward neural network, and an output layer. By alternating convolutional and pooling layers, the (raw) input (e.g., a RGB image) is progressively transformed into more abstract representations. Therefore, the convolutional layers convolve the input feature maps with a set of learnable filters (i.e., non-linear transformations) to produce a stack of output feature maps (Zeiler and Fergus, 2014). The pooling layers are used to reduce the dimensionality of the feature maps by computing some aggregation function (typically the maximum or the mean) across small local regions of the input (Boureau et al., 2010). This results in a hierarchical set of features where higher-level (more abstract) features are defined in terms of lower-level (less abstract) features (Najafabadi et al., 2015b).

The resulting feature maps are then concatenated and fed into a stack of fully connected neural layers to map these features onto the desired output.

167 Deep neural networks come with a lot of trainable parameters: order 100 million parameters is

168 not exceptional. In order to properly adjust the weights, gradient descent in combination with the

169 backpropagation procedure can be used (LeCun et al., 2015). By applying the chain rule on the

170 stacked layers on both the convolutional and fully connected layers, the gradient of the objective

171 with respect to the input can be computed. The backpropagation equation can be applied repeatedly

172 to propagate gradients through all modules, starting from the output at the top (where the network

173 produces its prediction) all the way to the bottom (where the external input is fed) (LeCun et al.,

174 2015). Today, adapted versions of the gradient descent optimization algorithm are used (see Ruder,

175 2016, for an overview). A particular popular optimizer is Adam (Kingma and Ba, 2014), an adaptive

176 learning-rate method, with bias-correction and momentum.

177 Because of their proven capabilities and state-of-the-art results in many domains, deep neural

178 networks are popular. However, due to their huge amount of trainable parameters, overfitting to data

179 remains a major challenge. A toolbox of techniques to avoid overfitting exist, including the reduction

180 of the model complexity by reducing the number of hidden layers or units, layer-wise pre-training and

181 fine-tuning (Bengio et al., 2007), dropout (Srivastava et al., 2014) and data augmentation (Simard

182 et al., 2003).

183 ### 2.2.2   Detection of stomata with deep learning

184 In this work, we assessed the performance of deep learning for the detection of stomata. While this

185 task can be broadened to a generic object detection task for which multiple efficient methodologies

186 were proposed (see Liu et al., 2018b, for an extensive review), we focus on a simple methodology

187 across multiple species which models the stomata detection task as classification task within a fixed

188 window. This baseline approach is illustrated in Figure 1.

189 For generating the training set we used herbarium specimens of 19 common tropical tree species
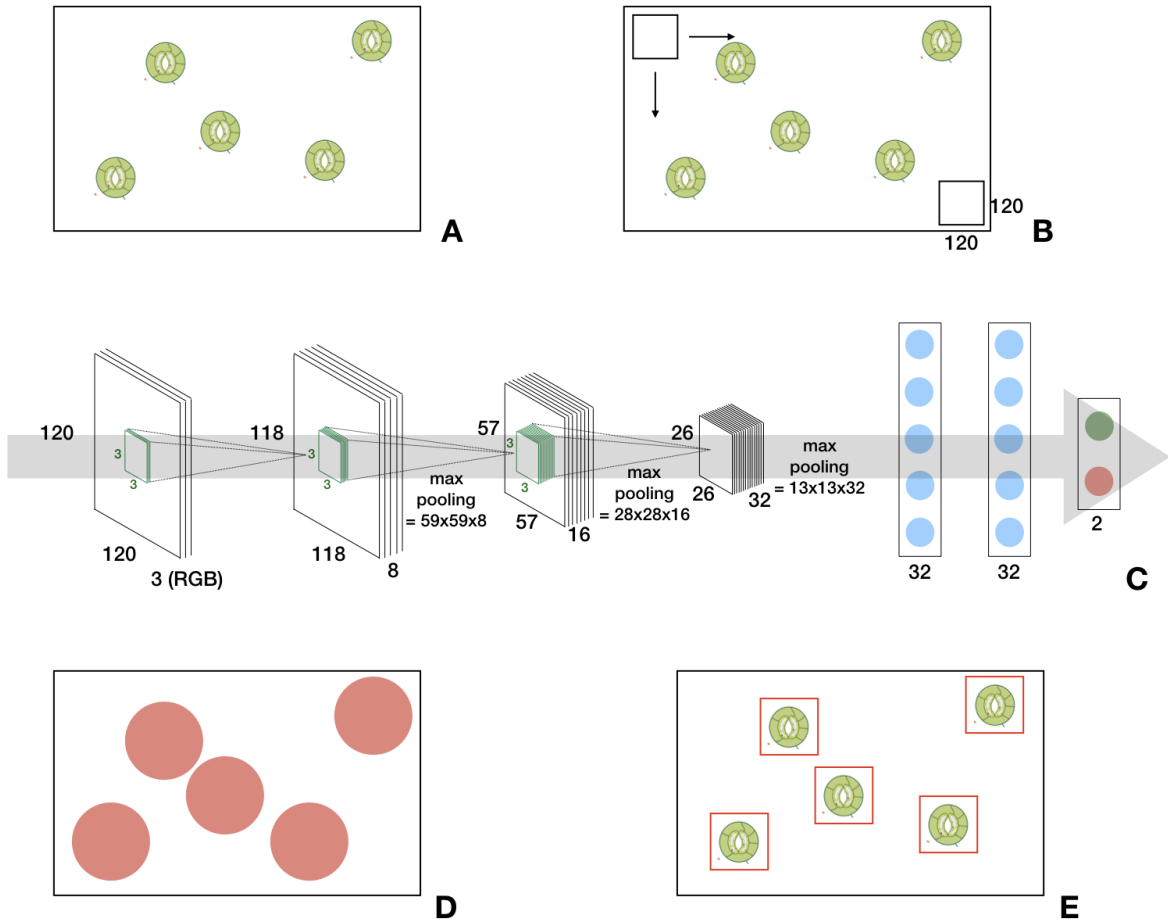
9

Figure 1: From leaf to label: a simple deep learning approach for automatic stomata detection. A photomicrograph (A) is divided into small overlapping patches (B) by using a sliding window approach. The deep learning architecture (C) is trained to label these patches. Positively labeled patches of a photomicrograph (D) are clustered which results in the detection (E).

190 belonging to 12 flowering plant families and 8 orders (Fig. 2, Suppl. Table). The choice of

191 trainingset was made in function of a running research project (COBECORE) to investigate the

192 change in stomatal density and function over time in Central-African tropical rainforest (Bauters

193 et al., under review). A total of 431 micrographs was used for training, 1-53 training images per
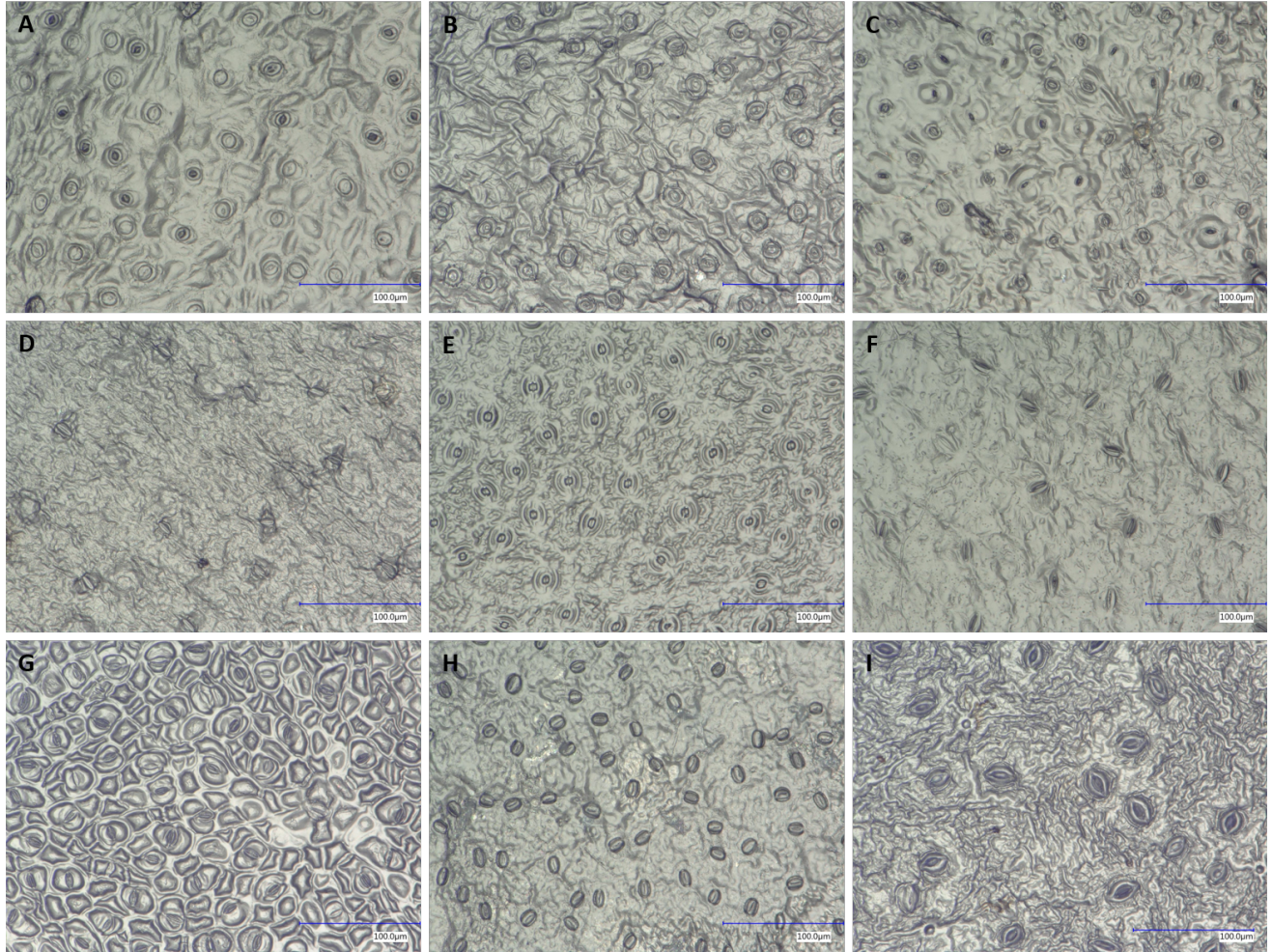
194 species, 3-115 per family, and 14-126 per order.

Figure 2: Stomata microscope images of herbarium specimens of nine representative species of the training set used to train the classification algorithm: *Cola griseiflora* (A), *Carapa procera* (B), *Celtis mildbraedii* (C), *Garcinia punctata* (D), *Mammea africana* (E), *Petersianthus macrocarpus* (F), *Prioria balsamifera* (G), *Erythrophleum suaveolens* (H), *Trichilia gigliana* (I).

195 In order to detect the stomata in a picture, we applied a simple patch-based method (Hou et al.,

196 2016; Cruz-Roa et al., 2014). Therefore, we divided each picture in multiple overlapping patches

197 of size 120 by 120 pixels. This patch size is based on the average stomatal size observed in the

training set. The patches were labeled as being positive or negative by an expert (Fig. 3). Note the variability of the stomata in the training set as well the variability of the negative patches due to the occurrence of different artifacts in the data. In total we extracted more than 12 thousand positive labeled patches and 72 thousand negative patches from the training set. Due to the apparent larger variability in the negative patches, more negative patches than positive patches were included.
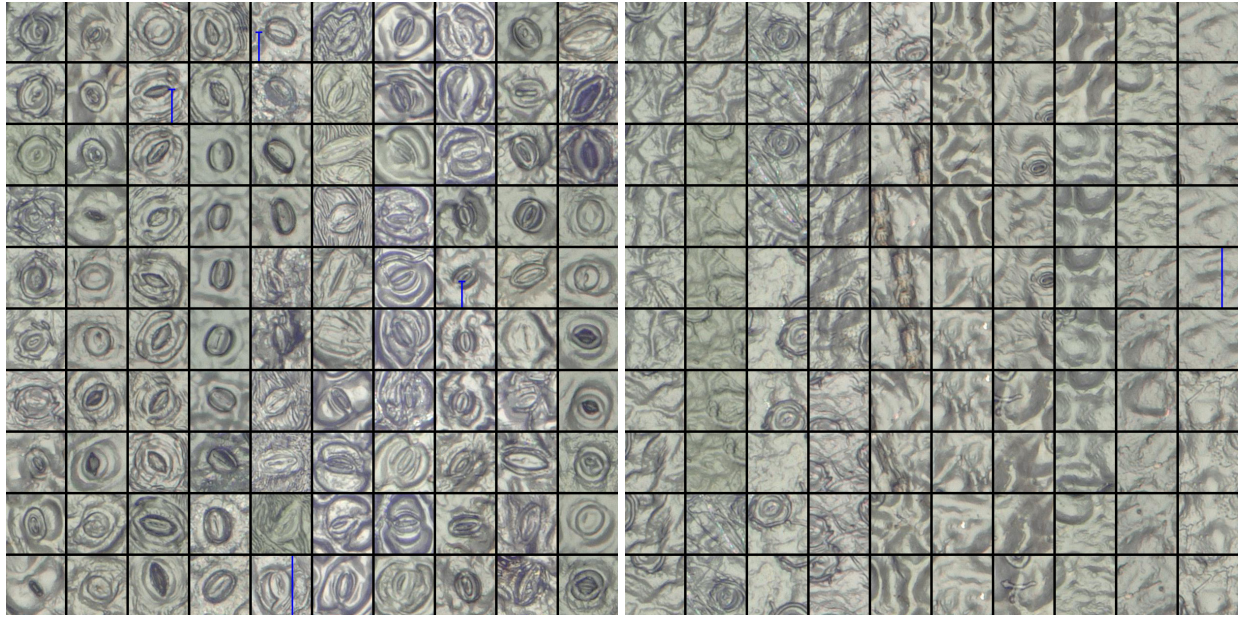

Figure 3: Patches with positive (left) and negative (right) examples of stomata. Stomata that are not fully visible, were labeled negative.

The obtained patches were then used to train three different deep learning models: two basic architectures with three convolutional layers followed by two dense layers and an output layer (Fig. 1C) with respectively 180,242 and 23,297,090 trainable parameters. Therefore we varied the depth of of the convolutional layers from 8-16-32 (basic shallow) and 32-64-128 (basic deep)), and the size of the dense layers: 2x32 neurons (basic shallow) and 2x1024 neurons (basic deep). One VGG19 (Simonyan and Zisserman, 2014) architecture with 47,297,602 parameters of which 27,273,218 were trained (i.e. the parameters from the fully connected layers) by fine tuning them

210 on our training set and 20,024,384 parameters (i.e. the parameters from the convolutional layers)

211 obtained through pre-training on ImageNet (Deng et al., 2009). These parameters were optimized

212 by using the Adam (Kingma and Ba, 2014) learning rule for which both the batch size and learning

213 rate were tuned. Dropout and data augmentation, by random rotations, horizontal and vertical flips

214 of the patches, were applied to avoid overfitting. Table 2.2.2 summarizes all the training parameters

215 of the deep learning architecture. Our deep learning models were trained (or fine-tuned in the case

216 of VGG19) over 200 epochs (50 epochs for VGG19) to output two numbers between 0.0 and 1.0

217 indicating the absence or presence of a stoma. Intuitively, the output is either $[0.0, 1.0]$ or $[1.0, 0.0]$

218 depending whether the patch contains the whole stoma or not. In reality, however, the network will

219 output any number between 0.0 and 1.0 depending on the model confidence. Consequently, one

220 has to tune a threshold with a validation set which, in this case, consisted of three plant species

221 belonging to the order of Sapindales (*Lannea acida, Lannea welwitschii* and *Lannea schweinfurthii*)

222 and are relatively closely related to the species from the training set (see Suppl. Table). Lastly,

223 all positively labeled patches are clustered by using mean shift clustering (Comaniciu and Meer,

224 2002). This technique groups neighboring (or even overlapping) positively labeled patches from

225 which the resulting stoma coordinates are derived. All software was implemented in Python 3.6.

226 Keras (Chollet et al., 2015) and Tensorflow (Abadi et al., 2016) were used to implement the deep

227 learning models. Training and testing was performed on a Linux (Ubuntu 18.04) workstation with

228 an i7-5930k CPU, 64 GB RAM and a Nvidia^TM Titan Xp GPU.

229   To evaluate the performance of the model, we calculated the information retrieval (IR) standard

230 measures, precision ($= \frac{TP}{TP+FP}$) and recall ($= \frac{TP}{TP+FN}$). Precision decreases with the number of false

231 positives (FP) and recall with the number of false negatives (FN). The F-score is the harmonic mean

232 of precision and recall with a high F-score, meaning low false positives and low false negatives.

233 Precision indices were calculated for all annotated images used for training (denoted "training set"

parameter	Basic shallow	Basic deep	VGG19
#parameters	180,242	23,297,090	47,297,602
#trainable parameters	180,242	23,297,090	27,273,218
optimizer	Adam	Adam	Adam
parameters optimizer	$\alpha = 5 \times 10^{-4}$	$\alpha = 5 \times 10^{-5}$	$\alpha = 5 \times 10^{-6}$
	$\beta_1 = 0.9$	$\beta_1 = 0.9$	$\beta_1 = 0.9$
	$\beta_2 = 0.999$	$\beta_2 = 0.999$	$\beta_2 = 0.999$
batch size	32	64	128
training epochs	200	200	50

Table 1: Summary of the training parameters.

in Suppl. Table), on 70 unseen images from a subset of the training set ("unseen within the scope of training") and on 595 images from species not included in the training set ("unseen beyond the scope of training"), a range of 16 species from 7 genera chosen from more and less related angiosperm orders as the samples used for training. The latter set was included to assess the performance of the model on other angiosperm species and to evaluate how well the model generalizes to these other species. We expected the deep learning model to perform better on species from the same angiosperm order as the training species as related species are expected to resemble each other more in stomatal shape and size (Zhang et al., 2012). As stomatal shape can vary in relation to climate even between species within a genus (e.g. Yukawa et al., 1992; Pautov et al., 2017) we sampled for this dataset three species within each genus with one species from the tropical rainforest, one from the tropical moist deciduous forest and one from tropical shrubland and desert (but only two climate regions for the Asparagaceae and one for Orchidaceae) to average precision measures and be able to compare genera by controlling for provenance. Precision indices for the training set were calculated to assess the performance in function of the number of stomata used per species for training and to compare performance to the "unseen beyond the scope of training" set.

The output of the developed model for stomatal detection consists of the coordinates of the

detected stomata. To calculate stomatal densities for scientific research questions, all stomata per image are counted and converted to the number of stomata per square millimeter. The accuracy (%) was calculated for 70 unseen images of species used in the training set ("unseen within the scope of training") for which we compared manual and computed stomatal counts. Accuracy is defined as the ratio of the number of correctly classified items to the total number of items (Michie et al., 1994). Goodness-of-fit was determined by calculation of the coefficient of determination (R^2) from a linear regression between computed and manual counts.

3   Results

3.1   Efficacy of the nail polish method

A total of 49 species was sampled from the African herbarium of Meise Botanic Garden (see Suppl. Table). The nail polish method was successfully applied in 78% of the species sampled. Generating impressions failed in 16 % of the species due to hairy or velvety leaf surfaces. In 7% of the species for which we managed to get leaf prints, we were unable to detect the stomata visually.

3.2   Model selection and evaluation

The accuracy of all three architectures on an unseen dataset (i.e. the validation set) is depicted in Fig. 4, illustrating the precision and recall for varying thresholds (0.05 to 0.95), as well as the F-score, which is an indication of the overall performance. One can observe that there is a trade-off between precision and recall, a well-known, general feature of information retrieval models. For example, one can choose to obtain maximal precision with very low recall or vice versa. From Fig. 4 it is clear that with increasing trainable parameters, the performance of the architecture increases,

although the VGG19 architecture only slightly outperforms the basic architectures. Furthermore, from Fig. 4 one can observe that the VGG19 architecture is less sensitive to the choice of the threshold in comparison to the basic architecture. Moreover, VGG19 can be seen as a standard textbook approach, while the basic architecture were hand-tuned. For all these reasons, we will continue our analysis and discussion with the VGG19 architecture. However, we want to point out that our choice is not the computationally most efficient. With less parameters, the basic architectures are less computationally demanding than VGG19. We refer to the work of Bianco et al. (2018) for a benchmark study of deep learning architectures.

Figure 5 shows that there are slight variations of the performance on the validation set. For the VGG19 architecture, a threshold equal to 0.7 is a good trade-off between precision and recall and will result in an average F-score of 0.89. This is close to an average F-score of 0.87 for the plant species of the training set. For species for which 250 stomata or more were used for training, precision, recall, and F-score values of 0.8 and higher were obtained (Fig. 6).

3.3   Accuracy

The accuracy was calculated for 70 images of species within the scope of the training set to compare results of stomatal densities between computed and manual counts. Average accuracy was high (94%) and a strong correlation between the computed counts and the manual counts was observed among all the images (Fig. 7, R^2 = 0.96, P < 0.001). Figure 7 shows the reference line (1:1) with an intercept within the 95% confidence interval (CI) around the intercept (-4.46-0.86) of the linear regression and with a slope value of 1 slightly outside the 95% CI of the regression slope (1.01-1.11). For images containing many stomata (>60) stomatal number tends to be underestimated (Fig. 7).
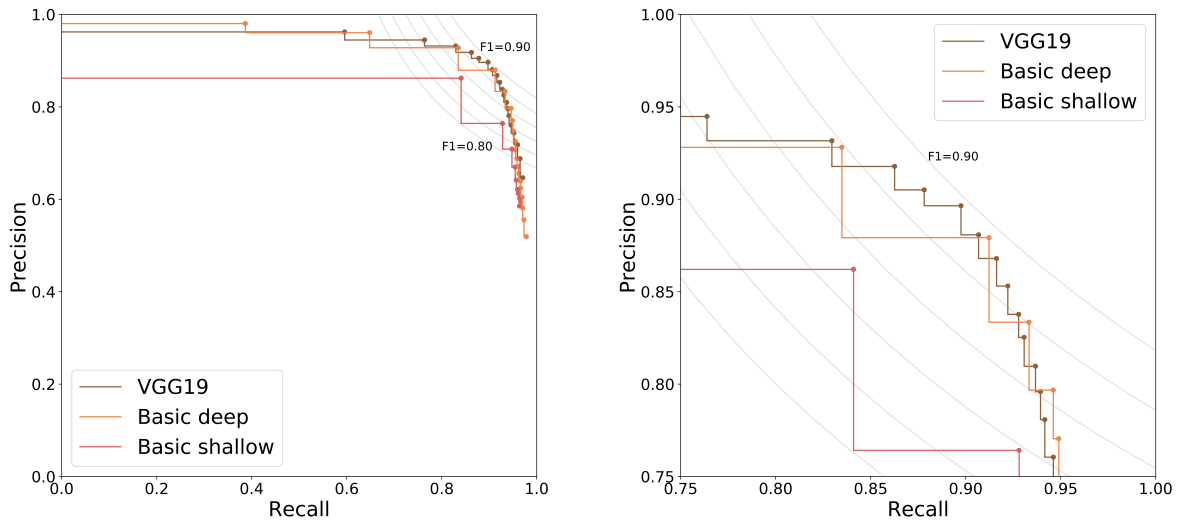
17

Figure 4: Precision-recall curve for the detection of three species unseen during training for three different deep learning architectures (zoom on the right). The curve gives us insight on how to choose the decision threshold which ranged from 0.05 to 0.95 in steps of 0.05. To guide this decision process the F iso−curves are shown as well.
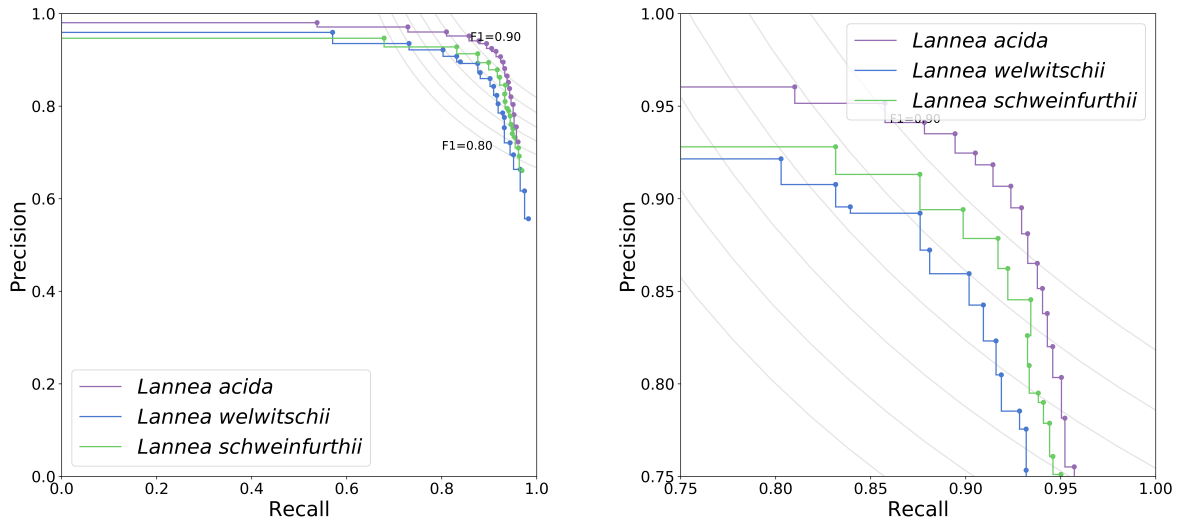
Figure 5: Precision-recall curve for the fine-tuned VGG19 architecture on the three different species (zoom on the right).

3.4   Generalisation to other species

In Figure 8 the overall performance of the VGG19 architecture on the "unseen beyond the scope of training set" (open circles) is shown for a confidence threshold of 0.7. Average precision, recall and F-score for the training set are 0.84, 0.91 and 0.87, respectively. Performance indices for unseen species within the same angiosperm order as the training set (Malpighiales, Ericales) range between 0.75-0.84 for precision, 0.57-0.87 for recall and 0.64-0.79 for F-score. Performance indices for unseen species beyond the training set (Poales, Asparagales, Gentianales, Solanales) range between 0.53-0.77 for precision, 0.63-0.94 for recall and 0.57-0.80 for F-score.
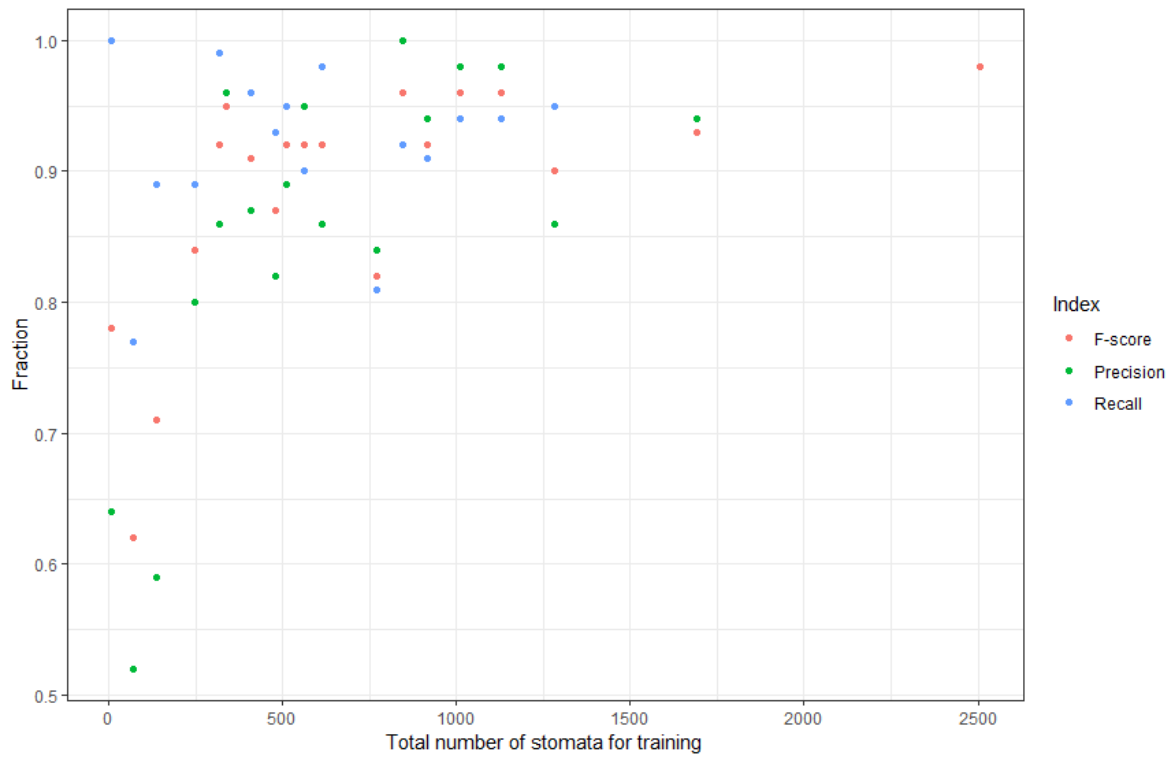
19

Figure 6: Precision, recall and F-score indices in function of the number of stomata used for each of the 19 species for training.
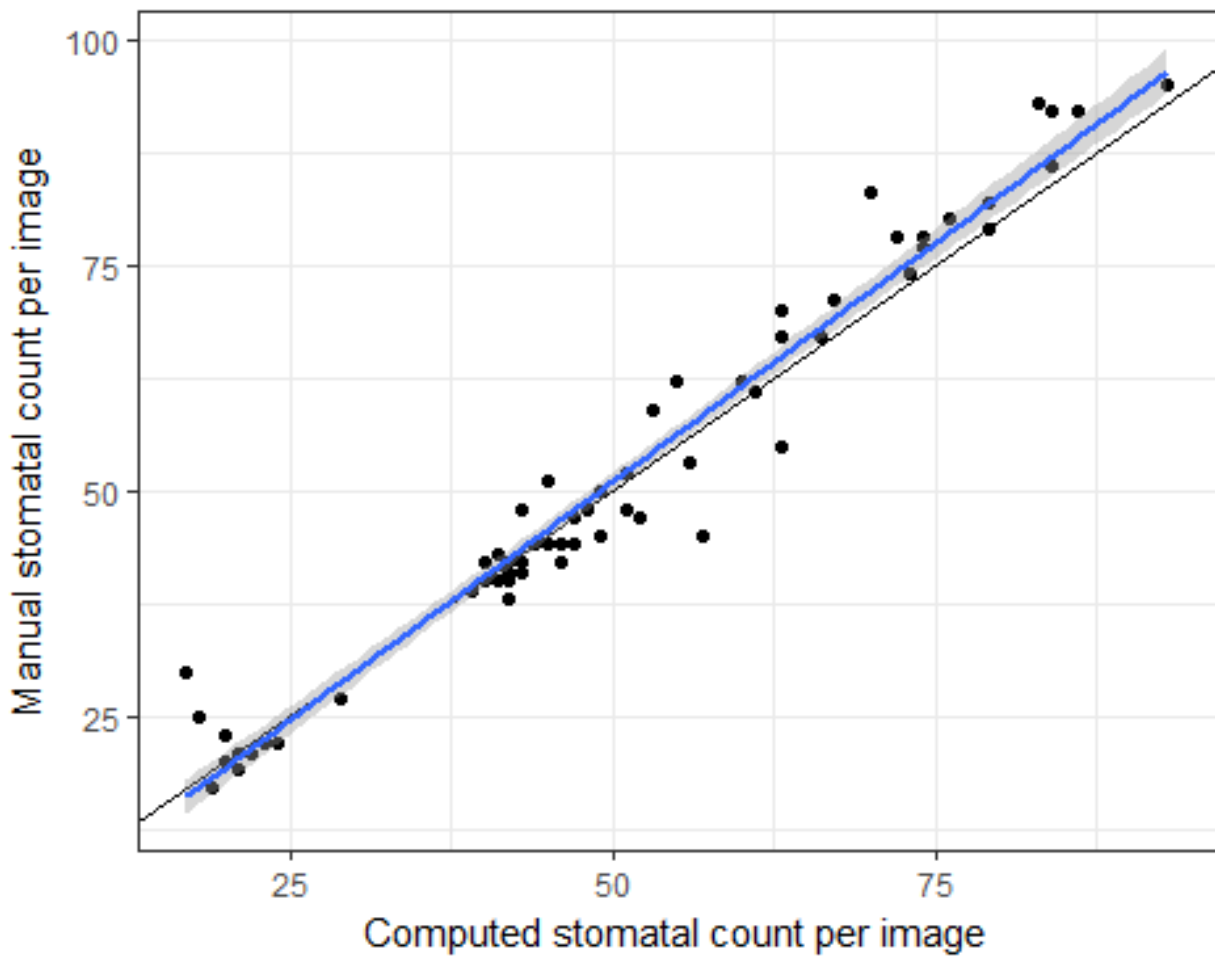
Figure 7: Accuracy of the computed stomatal counts per image ($n = 70$) from seven species included in the training set. The blue line with grey 95% CIs is the regression line with slope 1.056 and intercept -1.8 ($R^2 = 0.96$, $P < 0.001$)), the black line is the reference line (1:1).
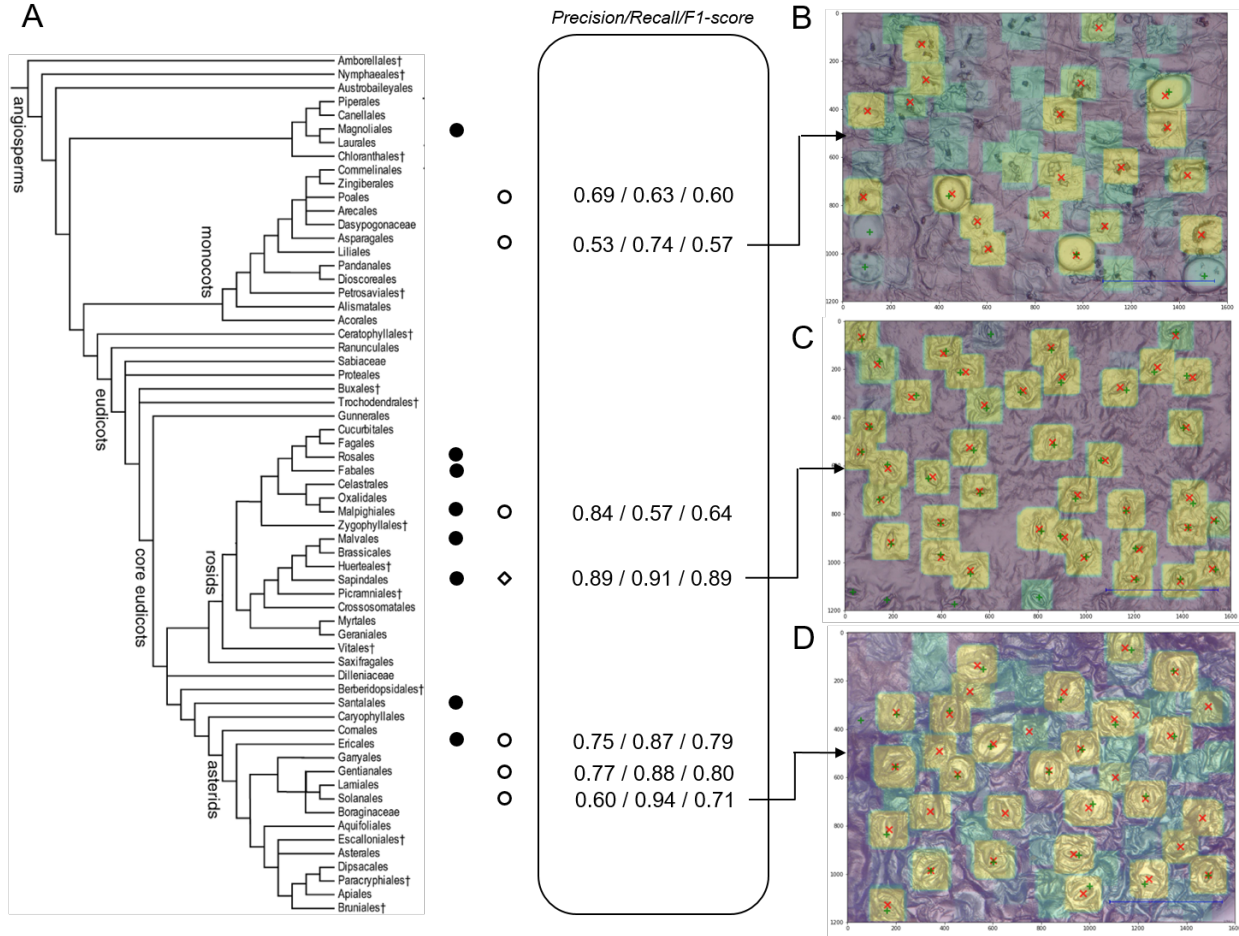
Figure 8: Performance of the network in function of the phylogenetic relatedness of taxa used for training, validation and testing. (A) Angiosperm phylogeny (modified from APG III, 2009) and diversity in the training set (full circles) and test set (open circles). The open diamond indicates the position within the angiosperm phylogeny of the taxa used for validation (*Lannea* species; see text). The numbers in the central frame denote the performance indices: precision, recall and F-score. Average precision, recall and F-score for the training set are 0.84, 0.91 and 0.87, respectively. Images (B, C, and D) visualize the performance of the network on unseen taxa belonging to the test set with (B) *Cyrtorchis chailluana* (Orchidaceae, Asparagales), (C) *Lannea schweinfurthii* (Anacardiaceae, Sapindales) and (D) *Ipomoea eriocarpa* (Convolvulaceae, Solanales). Green crosses denote the actual stomata, red x's the stomata recognized by the network with a confidence of 0.7 or higher. Color gradient from green (low confidence) to yellow (high confidence).

4 Discussion

In this work, we developed a leaf-to-label workflow that allows detecting stomata on light microscope images from dried plant material such as that of herbarium specimens. Even though mostly used in fresh plant material (e.g. Wu and Zhao, 2017), the nail polish method proves to be a reliable, non-invasive, easy, and inexpensive method that can obtain qualitative leaf impressions from dried leaves on the majority of species (78%). We trained a deep learning architecture for the detection of stomata in focus-stacked images of high-resolution. However, we believe that traditional light microscopy could also be used for imaging given that the entire field of view is in focus. We illustrated that, even with a simple deep learning approach in which we model the object detection problem as a classification problem with a fixed patch size based, a F-score of 0.89 can be reached on unseen taxa on the condition that they are in the phylogenetic scope of the training set. This is in line with the average results (F-score:0.87) on the training set. The model on average did not perform better on unseen species within the same angiosperm order as the training set (F-score: 0.64-0.79) as compared to its performance on unseen species of other angiosperm orders (F-score: 0.57-0.80). This result seems to indicate that the variation in stomatal structure and shape within flowering plant orders is similar to the variation between them. However, note that this test set includes at most a few species representing an angiosperm order and therefore does not include all variation within genera, families, and orders. The training focused mainly on taxa belonging to the core eudicots and one species of the basal angiosperms *Polyalthia suaveolens* (Magnoliales). The model performed on average better on unseen species from the core eudicots (F-score: 0.77) than on unseen species from the monocots (F-score: 0.59). The difference in stomatal shape between monocots and the dicots is apparent, especially the grasses (Poaceae), represented here by *Chloris* species are known for their particular dumbbell-shaped guard cells as compared to kidney-shaped cells of dicots (Zeiger et al.,

322 1987; Rudall et al., 2017). Also, the orchid species included in the test set, *Cyrtorchis chailluana*

323 has a stomatal shape not easily detected by our model (Fig. 8), probably because of its particular

324 circular shape and round opening [*cf.* stoma Type II in *Dendrobium* (Yukawa et al., 1992)]. Also

325 note that we did not include species with extremely large stomata typical for e.g. the Liliaceae, as

326 this will decrease the performance of the model to detect the stomata (but see below).

327   While the model performs relatively well over a broad taxonomy, our approach has room for

328 further improvement. Firstly, the model performance is highly related to the variation (Fig. 8), the

329 quantity (Fig. 6), and quality of training images. The network presented in this paper is not trained

330 to handle low-quality images. Therefore high-quality images should be aimed to enable the network

331 to perform optimally. The quality (contrast, blurriness, etc.) of a set of pictures can be quantified

332 using the image histogram and using PyImageQualityRanking software for ranking the images in a

333 set and detecting outliers (Koho et al., 2016). Based on this ranking, one can choose to leave out

334 low-ranked images due to their insufficient quality. If low-quality images should still be processed,

335 then the network should be trained accordingly.

336   Secondly, the performance of our model depends on the number of examples of stomata used

337 during training (Fig. 6). In general, if more examples are available of a species, the better the

338 performance of the model for that species. In this case, even though overall accuracy of stomatal

339 counts was very high (94%) for unseen pictures of seven species (Fig. 7), the average accuracy for

340 each of the species individually was highly correlated with the number of images and total number

341 of stomata seen during training (results not shown) as was the case for the information retrieval

342 (IR) standard measures (Fig. 6). For example, only 69 stomata or 3 images for the species *Irvingia*

343 *grandifolia* were used in training (Suppl. Table) rendering an average accuracy of 64%, i.e. a

344 reduction or increase in stomatal density of 36%. Since a 28% reduction in stomatal density in

345 transgenic poplars is enough to cause a 30% drop in transpiration (Wang et al., 2016; Bertolino

24

346   et al., 2019) we recommend at least 250 stomata for training depending on the level of difference in

347   stomatal density one wants to detect. If small differences in stomatal density within a species are

348   targeted, the general protocol described in this paper can be used. In order to obtain more accurate

349   results, the threshold of the deep learning model (cf. Section 3.2) can be adjusted for each species

350   separately. Furthermore, the accuracy for an individual species can be increased by fine-tuning the

351   model by training the dense layers of the deep learning model.

352   Thirdly, our patch-based approach is constrained by a patch size of 120 by 120 pixels which

353   correspond to a window of 25 by 25 μm using the microscope settings as described above.

354   Angiosperms on average have a stomatal length or guard cell length of 31 μm (cf. Hodgson et al.,

355   2010; Beaulieu et al., 2008) (Suppl. Fig.). Although this patch size could be successfully applied to

356   the majority of angiosperm species, the patch size limits both the aspect ratio and the scale of the

357   input image. The simplest solution is to adjust magnification during data collection, by increasing

358   the magnification when stomata are too small to be detected and decreasing the magnification when

359   stomata extend beyond a patch size of 120 by 120 pixels. In our model, stomata between 60 and 120

360   pixels are best detected by the model. Another more elegant way of handling this problem is by

361   including some region of interest pooling layer as discussed by Dai et al. (2016) and He et al. (2014),

362   which would allow moving from the patch-based method to detect all stomata in a spatial hierarchical

363   way. This object detection pipeline can be improved further with Fast R-CNN (Girshick, 2015) and

364   Faster R-CNN (Ren et al., 2015) which combine the idea of using a spatial hierarchical pooling

365   with region based convolutions into an end-to-end trainable deep learning model. Furthermore, if

366   processing speed is an issue, one can opt for a single shot multibox detector approach (SSD). SSD,

367   discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios

368   and scales per feature map location (Liu et al., 2016). At prediction time, the network generates

369   scores for the presence of each object category in each default box and produces adjustments to the

370   box to better match the object shape.

371   Fourthly, with the current advances in deep learning, the object detection pipeline can be improved

372   further by using novel convolutional neural network architectures such as Xception (Chollet, 2017)

373   or ResNeXt (Xie et al., 2017) as a backbone for feature extraction. See (Bianco et al., 2018) for an

374   in-depth analysis of the majority of the deep neural network architectures that deviate from the idea

375   that simply stacking convolutional layers is sufficient.

376   To summarize, we illustrated that by using a simple deep learning architecture one can work out

377   a simple leaf-to-label workflow that allows detecting stomata on light microscope images from dried

378   plant material such as that of herbarium specimens. Our approach can be optimized depending

379   on the availability of the data as well as by using more recent object detection pipelines. We

380   recommend the survey paper of Liu et al. (2018b) and (Huang et al., 2017) for a thorough overview

381   and benchmarking of object detector pipelines.

5   Conclusions

383   The entire leaf-to-label pipeline presented here could be of use in different research areas with the

384   need for stomatal count data of many specimens. It will allow ecologists to focus on the ecological

385   questions rather than on the technical aspects of data analysis and more specifically deep learning,

386   and computer scientists to pave new roads on some of the biological world's most complex units,

387   such as ecosystems (Christin et al., 2019). Large-scale studies using stomata of fossils to reconstruct

388   a changing environment in deep time (e.g. McElwain et al., 1999; Franks et al., 2017)), as well as

389   work on the anthropogenic effect on stomatal density and size in agricultural crops (Zheng et al.,

390   2013) could benefit from such an approach i.e. the use of a general deep learning model that can

391   be tweaked and expanded for the detection of other objects such as epidermal cells. Especially

the information locked in the archives of herbaria, the result of century-long efforts of collecting, has shown to be of great value in several studies, as the digitization of herbaria specimens has the potential to produce data to facilitate the study of the natural world (Goodwin et al., 2015). The leaf-to-image approach described here, is easy to perform and, given that imaging technology is becoming faster and can be partially automated, the exploration of these sleeping beauties is within reach.

Acknowledgements

Data and Material Accessibility

The trained model is accessible to use at https://kiks.ilabt.imec.be/jupyter/ in Jupyter Notebook with a user-friendly interface and guidelines where individual microscope images can be uploaded. All light microscope images used in this study are made freely accessible on Zenodo under the CC-by license (DOI 10.5281/zenodo.3579227). The herbarium specimens as referred to in the

413   Supplementary Table can be visualized in the virtual herbarium of Meise Botanic Garden following

414   this link: http://www.botanicalcollections.be/specimen/*barcode*.

References

Abadi, Martín, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. 2016. "Tensorflow: Large-scale machine learning on heterogeneous distributed systems." *arXiv preprint arXiv:1603.04467* .

Aono, Alexandre, James Nagai, Gabriella Dickel, Rafaela Marinho, Paulo Oliveira, and Fabio Faria. 2019. "A Stomata Classification and Detection System in Microscope Images of Maize Cultivars." *bioRxiv* .

Beaulieu, Jeremy M., Ilia J. Leitch, Sunil Patel, Arjun Pendharkar, and Charles A. Knight. 2008. "Genome size is a strong predictor of cell size and stomatal density in angiosperms." *New Phytologist* 179:975–986.

Bengio, Yoshua, Pascal Lamblin, Dan Popovici, and Hugo Larochelle. 2007. "Greedy layer-wise training of deep networks." In *Advances in neural information processing systems*, pp. 153–160.

Berry, J A, D J Beerling, and P J Franks. 2010. "Stomata: key players in the earth system, past and present." *Current Opinion in Plant Biology* 13:233–240.

Bertolino, Lígia T., Robert S. Caine, and Julie E. Gray. 2019. "Impact of Stomatal Density and Morphology on Water-Use Efficiency in a Changing World." *Frontiers in Plant Science* 10:225.

Bianco, Simone, Remi Cadene, Luigi Celona, and Paolo Napoletano. 2018. "Benchmark analysis of representative deep neural network architectures." *IEEE Access* 6:64270–64277.

Boureau, Y-Lan, Jean Ponce, and Yann LeCun. 2010. "A theoretical analysis of feature pooling in visual recognition." In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 111–118.

Butler, Ethan E., Abhirup Datta, Habacuc Flores-Moreno, Ming Chen, Kirk R. Wythers, Farideh Fazayeli, Arindam Banerjee, Owen K. Atkin, Jens Kattge, Bernard Amiaud, Benjamin Blonder, Gerhard Boenisch, Ben Bond-Lamberty, Kerry A. Brown, Chaeho Byun, Giandiego Campetella, Bruno E. L. Cerabolini, Johannes H. C. Cornelissen, Joseph M. Craine, Dylan Craven, Franciska T. de Vries, Sandra Díaz, Tomas F. Domingues, Estelle Forey, Andrés González-Melo, Nicolas Gross, Wenxuan Han, Wesley N. Hattingh, Thomas Hickler, Steven Jansen, Koen Kramer, Nathan J. B. Kraft, Hiroko Kurokawa, Daniel C. Laughlin, Patrick Meir, Vanessa Minden, Ülo Niinemets, Yusuke Onoda, Josep Peñuelas, Quentin Read, Lawren Sack, Brandon Schamp, Nadejda A. Soudzilovskaia, Marko J. Spasojevic, Enio Sosinski, Peter E. Thornton, Fernando Valladares, Peter M. van Bodegom, Mathew Williams, Christian Wirth, and Peter B. Reich. 2017. "Mapping local and global variability in plant trait distributions." *Proceedings of the National Academy of Sciences* 114:E10937–E10946.

Chollet, François. 2017. "Xception: Deep learning with depthwise separable convolutions." *arXiv preprint* pp. 1610–02357.

Chollet, François et al. 2015. "Keras." `https://keras.io`.

Christin, Sylvain, Éric Hervet, and Nicolas Lecomte. 2019. "Applications for deep learning in ecology." *Methods in Ecology and Evolution* 10:1632–1644.

Comaniciu, Dorin and Peter Meer. 2002. "Mean shift: A robust approach toward feature space analysis." *IEEE Transactions on pattern analysis and machine intelligence* 24:603–619.

Cornelissen, JHC, S Lavorel, E Garnier, S Diaz, N Buchmann, DE Gurvich, PB Reich, H Ter Steege, HD Morgan, MGA Van Der Heijden, et al. 2003. "A handbook of protocols for standardised and easy measurement of plant functional traits worldwide." *Australian journal of Botany* 51:335–380.

Cruz-Roa, Angel, Ajay Basavanhally, Fabio González, Hannah Gilmore, Michael Feldman, Shridar Ganesan, Natalie Shih, John Tomaszewski, and Anant Madabhushi. 2014. "Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks." In *Medical Imaging 2014: Digital Pathology*, volume 9041, p. 904103. International Society for Optics and Photonics.

da Silva Oliveira, Marcos William, Núbia Rosa da Silva, Dalcimar Casanova, Luiz Felipe Souza Pinheiro, Rosana Marta Kolb, and Odemir Martinez Bruno. 2014. "Automatic Counting of Stomata in Epidermis Microscopic Images." .

Dai, Jifeng, Yi Li, Kaiming He, and Jian Sun. 2016. "R-fcn: Object detection via region-based fully convolutional networks." In *Advances in neural information processing systems*, pp. 379–387.

Dawson, Samantha Katherine, Lynne Boddy, Hans Halbwachs, Claus Bässler, Carrie Andrew, Thomas Ward Crowther, Jacob Heilmann-Clausen, Jenni Nordén, Otso Ovaskainen, and Mari Jönsson. 2019. "Handbook for the measurement of macrofungal functional traits: A start with basidiomycete wood fungi." *Functional Ecology* 33:372–387.

Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. "Imagenet: A large-scale hierarchical image database." In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 248–255. Ieee.

Dieleman, Sander, Kyle W Willett, and Joni Dambre. 2015. "Rotation-invariant convolutional neural networks for galaxy morphology prediction." *Monthly notices of the royal astronomical society* 450:1441–1459.

Drake, Paul L, Ray H Froend, and Peter J Franks. 2013. "Smaller, faster stomata: scaling of stomatal size, rate of response, and stomatal conductance." *Journal of Experimental Botany* 64:495–505.

Duarte, Kauê TN, Marco António Garcia de Carvalho, and Paulo S Martins. 2017. "Segmenting High-quality Digital Images of Stomata using the Wavelet Spot Detection and the Watershed Transform." In *VISIGRAPP (4: VISAPP)*, pp. 540–547.

Fetter, Karl, Sven Eberhardt, Rich S Barclay, Scott Wing, and Stephen R Keller. 2018. "StomataCounter: a deep learning method applied to automatic stomatal identification and counting." *bioRxiv* p. 327494.

Franks, Peter J, Joseph A Berry, Danica L Lombardozzi, and Gordon B Bonan. 2017. "Stomatal function across temporal and spatial scales: deep-time trends, land-atmosphere coupling and global models." *Plant Physiology* 174:583–602.

Girshick, Ross. 2015. "Fast r-cnn." In *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448.

Goodwin, Zoë A, David J Harris, Denis Filer, John RI Wood, and Robert W Scotland. 2015. "Widespread mistaken identity in tropical plant collections." *Current Biology* 25:R1066–R1067.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2014. "Spatial pyramid pooling in deep convolutional networks for visual recognition." In *European conference on computer vision*, pp. 346–361. Springer.

Hetherington, Alistair M and F Ian Woodward. 2003. "The role of stomata in sensing and driving environmental change." *Nature* 424:901.

Higaki, Takumi, Natsumaro Kutsuna, and Seiichiro Hasezawa. 2014. "CARTA-based semi-automatic detection of stomatal regions on an Arabidopsis cotyledon surface." *PLANT MORPHOLOGY* 26:9–12.

Hodgson, J. G., M. Sharafi, A. Jalili, S. Díaz, G. Montserrat-Martí, C. Palmer, B. Cerabolini, S. Pierce, B. Hamzehee, Y. Asri, Z. Jamzad, P. Wilson, J. A. Raven, S. R. Band, S. Basconcelo, A. Bogard, G. Carter, M. Charles, P. Castro-Díez, J. H. C. Cornelissen, G. Funes, G. Jones, M. Khoshnevis, N. Pérez-Harguindeguy, M. C. Pérez-Rontomé, F. A. Shirvany, F. Vendramini, S. Yazdani, R. Abbas-Azimi, S. Boustani, M. Dehghan, J. Guerrero-Campo, A. Hynd, E. Kowsary, F. Kazemi-Saeed, B. Siavash, P. Villar-Salvador, R. Craigie, A. Naqinezhad, A. Romo-Díez, L. de Torres Espuny, and E. Simmons. 2010. "Stomatal vs. genome size in angiosperms: the somatic tail wagging the genomic dog?" *Annals of Botany* 105:573–584.

Hoo-Chang, Shin, Holger R Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura, and Ronald M Summers. 2016. "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning." *IEEE transactions on medical imaging* 35:1285.

31

Hou, Le, Dimitris Samaras, Tahsin M Kurc, Yi Gao, James E Davis, and Joel H Saltz. 2016. "Patch-based convolutional neural network for whole slide tissue image classification." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2424–2433.

Huang, Jonathan, Vivek Rathod, Chen Sun, Menglong Zhu, Anoop Korattikara, Alireza Fathi, Ian Fischer, Zbigniew Wojna, Yang Song, Sergio Guadarrama, et al. 2017. "Speed/accuracy trade-offs for modern convolutional object detectors." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7310–7311.

Jackson, Aaron S., Adrian Bulat, Georgios Tzimiropoulos, Michael P. Pound, Tony P. Pridmore, Andrew P. French, Alexandra J. Townsend, Darren M. Wells, Erik H. Murchie, Jonathan A. Atkinson, Marcus Griffiths, and Michael H. Wilson. 2017. "Deep machine learning provides state-of-the-art performance in image-based plant phenotyping." *GigaScience* 6.

Jayakody, Hiranya, Scarlett Liu, Mark Whitty, and Paul Petrie. 2017. "Microscope image based fully automated stomata detection and pore measurement method for grapevines." *Plant methods* 13:94.

Jefferson, Jennifer L, Reed M Maxwell, and Paul G Constantine. 2017. "Exploring the sensitivity of photosynthesis and stomatal resistance parameters in a land surface model." *Journal of Hydrometeorology* 18:897–915.

Kala, Jatin, Martin G De Kauwe, Andy J Pitman, Belinda E Medlyn, Ying-Ping Wang, Ruth Lorenz, and Sarah E Perkins-Kirkpatrick. 2016. "Impact of the representation of stomatal conductance on model projections of heatwave intensity." *Scientific reports* 6:23418.

Kingma, Diederik P and Jimmy Ba. 2014. "Adam: A method for stochastic optimization." *arXiv preprint arXiv:1412.6980* .

Koho, Sami, Elnaz Fazeli, John E. Eriksson, and Pekka E. Hänninen. 2016. "Image Quality Ranking Method for Microscopy." *Scientific Reports* 6:28962 EP –. Article.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton. 2012. "Imagenet classification with deep convolutional neural networks." In *Advances in neural information processing systems*, pp. 1097–1105.

Kushnir, Yochanan, Adam A Scaife, Raymond Arritt, Gianpaolo Balsamo, George Boer, Francisco Doblas-Reyes, Ed Hawkins, Masahide Kimoto, Rupa Kumar Kolli, Arun Kumar, et al. 2019. "Towards operational predictions of the near-term climate." *Nature Climate Change* p. 1.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. "Deep learning." *nature* 521:436.

Lin, Yan-Shih, Belinda E Medlyn, Remko A Duursma, I Colin Prentice, Han Wang, Sofia Baig, Derek Eamus, Victor Resco de Dios, Patrick Mitchell, David S Ellsworth, et al. 2015. "Optimal stomatal behaviour around the world." *Nature Climate Change* 5:459.

Liu, Congcong, Nianpeng He, Jiahui Zhang, Ying Li, Qiufeng Wang, Lawren Sack, and Guirui Yu. 2018a. "Variation of stomatal traits from cold temperate to tropical forests and association with water use efficiency." *Functional ecology* 32:20–28.

Liu, Li, Wanli Ouyang, Xiaogang Wang, Paul Fieguth, Jie Chen, Xinwang Liu, and Matti Pietikäinen. 2018b. "Deep learning for generic object detection: A survey." *arXiv preprint arXiv:1809.02165* .

Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. 2016. "Ssd: Single shot multibox detector." In *European conference on computer vision*, pp. 21–37. Springer.

McElwain, JC, DJ Beerling, and FI Woodward. 1999. "Fossil plants and global warming at the Triassic-Jurassic boundary." *Science* 285:1386–1390.

Michie, Donald, D. J. Spiegelhalter, C. C. Taylor, and John Campbell (eds.). 1994. *Machine Learning, Neural and Statistical Classification*. Upper Saddle River, NJ, USA: Ellis Horwood.

Moretti, Marco, André T. C. Dias, Francesco de Bello, Florian Altermatt, Steven L. Chown, Francisco M. Azcárate, James R. Bell, Bertrand Fournier, Mickaël Hedde, Joaquín Hortal, Sébastien Ibanez, Erik Öckinger, José Paulo Sousa, Jacintha Ellers, and Matty P. Berg. 2017. "Handbook of protocols for standardized measurement of terrestrial invertebrate functional traits." *Functional Ecology* 31:558–567.

Najafabadi, Maryam M., Flavio Villanustre, Taghi M. Khoshgoftaar, Naeem Seliya, Randall Wald, and Edin Muharemagic. 2015a. "Deep learning applications and challenges in big data analytics." *Journal of Big Data* 2:1.

Najafabadi, Maryam M, Flavio Villanustre, Taghi M Khoshgoftaar, Naeem Seliya, Randall Wald, and Edin Muharemagic. 2015b. "Deep learning applications and challenges in big data analytics." *Journal of Big Data* 2:1.

Parkhi, Omkar M, Andrea Vedaldi, Andrew Zisserman, et al. 2015. "Deep face recognition." In *BMVC*, volume 1, p. 6.

Pautov, Anatoly, Svetlana Bauer, Olga Ivanova, Elena Krylova, Yulia Sapach, and Galina Gussarova. 2017. "Role of the outer stomatal ledges in the mechanics of guard cell movements." *Trees* 31:125–135.

Perez-Harguindeguy, N., S. Diaz, E. Garnier, S. Lavorel, H. Poorter, P. Jaureguiberry, M.S. Bret-Harte, W.K. Cornwell, J.M. Craine, D.E. Gurvich, C. Urcelay, E.J. Veneklaas, P.B. Reich, L. Poorter, I.J. Wright, P. Ray, L. Enrico, J.G. Pausas, A.C. de Vos, N. Buchmann, G. Funes, F. Quetier, J.G. Hodgson, K. Thompson, H.D. Morgan, H. ter Steege, M.G.A. van der Heijden, L. Sack, B. Blonder, P. Poschlod, M.V. Vaieretti, G. Conti, A.C. Staver, S. Aquino, and J.H.C. Cornelissen. 2013. "New handbook for standardised measurement of plant functional traits worldwide." *Australian Journal of Botany* 61:167–234.

Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. 2015. "Faster r-cnn: Towards real-time object detection with region proposal networks." In *Advances in neural information processing systems*, pp. 91–99.

Rudall, Paula J., Elisabeth D. Chen, and Erin Cullen. 2017. "Evolution and development of monocot stomata." *American Journal of Botany* 104:1122–1141.

Ruder, Sebastian. 2016. "An overview of gradient descent optimization algorithms." *arXiv preprint arXiv:1609.04747* .

Scarlett, LIU, Julie Tang, Paul Petrie, and Mark Whitty. 2016. "A Fast Method to Measure Stomatal Aperture by MSER on Smart Mobile Phone." In *Applied Industrial Optics: Spectroscopy, Imaging and Metrology*, pp. AIW2B–2. Optical Society of America.

Shen, Dinggang, Guorong Wu, and Heung-Il Suk. 2017. "Deep learning in medical image analysis." *Annual review of biomedical engineering* 19:221–248.

Simard, Patrice Y, Dave Steinkraus, and John C Platt. 2003. "Best practices for convolutional neural networks applied to visual document analysis." In *null*, p. 958. IEEE.

Simonyan, Karen and Andrew Zisserman. 2014. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* .

Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. "Dropout: a simple way to prevent neural networks from overfitting." *The Journal of Machine Learning Research* 15:1929–1958.

Steinthorsdottir, Margret, F Ian Woodward, Finn Surlyk, and Jennifer C McElwain. 2012. "Deep-time evidence of a link between elevated CO2 concentrations and perturbations in the hydrological cycle via drop in plant transpiration." *Geology* 40:815–818.

Tanaka, Yu, Shigeo S. Sugano, Tomoo Shimada, and Ikuko Hara-Nishimura. 2013. "Enhancement of leaf photosynthetic capacity through increased stomatal density in Arabidopsis." *New Phytologist* 198:757–764.

Vialet-Chabrand, Silvère and Oliver Brendel. 2014. "Automatic measurement of stomatal density from microphotographs." *Trees* 28:1859–1865.

Wäldchen, Jana and Patrick Mäder. 2018. "Machine learning for image based species identification." *Methods in Ecology and Evolution* 9:2216–2225.

Wang, Congpeng, Sha Liu, Yan Dong, Ying Zhao, Anke Geng, Xinli Xia, and Weilun Yin. 2016. "PdEPF1 regulates water-use efficiency and drought tolerance by modulating stomatal density in poplar." *Plant Biotechnology Journal* 14:849–860.

Wang, Ruili, Guirui Yu, Nianpeng He, Qiufeng Wang, Ning Zhao, Zhiwei Xu, and Jianping Ge. 2015. "Latitudinal variation of leaf stomatal traits from species to community level in forests: linkage with ecosystem productivity." In *Scientific reports*.

Wolz, Kevin, Timothy Wertin, Mark Abordo, Dan Wang, and Andrew D. B. Leakey. 2017. "Diversity in stomatal function is integral to modelling plant carbon and water fluxes." *Nature Ecology Evolution* 1.

Woodward, F. I. 1987. "Stomatal numbers are sensitive to increases in CO2 from pre-industrial levels." *Nature* 327:617–618.

Wu, Shuchi and Bingyu Zhao. 2017. *Using Clear Nail Polish to Make Arabidopsis Epidermal Impressions for Measuring the Change of Stomatal Aperture Size in Immune Response*, volume 1578, pp. 243–248.

Xie, Saining, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. 2017. "Aggregated residual transformations for deep neural networks." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1492–1500.

Yukawa, Tomohisa, Toshio Ando, Kohji Karasawa, and Kiyoshi Hashimoto. 1992. "Existence of Two Stomatal Shapes in the Genus Dendrobium (Orchidaceae) and Its Systematic Significance." *American Journal of Botany* 79:946–952.

Zeiger, E., G.D. Farquhar, and I.R. Cowan. 1987. *Stomatal Function*. Stanford University Press.

Zeiler, Matthew D and Rob Fergus. 2014. "Visualizing and understanding convolutional networks." In *European conference on computer vision*, pp. 818–833. Springer.

Zhang, Shi-Bao, Zhi-Jie Guan, Mei Sun, Juan-Juan Zhang, Kun-Fang Cao, and Hong Hu. 2012. "Evolutionary Association of Stomatal Traits with Leaf Vein Density in Paphiopedilum, Orchidaceae." *PloS one* 7:e40080.

Zheng, Yunpu, Ming Xu, Ruixing Hou, Ruichang Shen, Shuai Qiu, and Zhu Ouyang. 2013. "Effects of experimental warming on stomatal traits in leaves of maize (Z ea may L.)." *Ecology and Evolution* 3:3095–3111.

640   WORD COUNT