Ecological Informatics 40 (2017) 50–56

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

LeafNet: A computer vision system for automatic plant species identification

Pierre Barré[a,*], Ben C. Stöver[b], Kai F. Müller[b], & Volker Steinhage[a]

[a] University of Bonn, Institute of Computer Science IV, Friedrich-Ebert-Allee 144, 53113 Bonn, Germany
[b] Westfälische Wilhelms-Universität Münster, Institute for Evolution and Biodiversity and Botanical Garden, Hüfferstrasse 1, 48149 Münster, Germany

ARTICLE INFO

Keywords:
Plant classification
Feature representation
Deep learning
Convolutional neural network
Convolutional layers
Feature maps

ABSTRACT

Aims: Taxon identification is an important step in many plant ecological studies. Its efficiency and reproducibility might greatly benefit from partly automating this task. Image-based identification systems exist, but mostly rely on hand-crafted algorithms to extract sets of features chosen a priori to identify species of selected taxa. In consequence, such systems are restricted to these taxa and additionally require involving experts that provide taxonomical knowledge for developing such customized systems. The aim of this study was to develop a deep learning system to learn discriminative features from leaf images along with a classifier for species identification of plants. By comparing our results with customized systems like *LeafSnap* we can show that learning the features by a convolutional neural network (CNN) can provide better feature representation for leaf images compared to hand-crafted features.

Methods: We developed *LeafNet*, a CNN-based plant identification system. For evaluation, we utilized the publicly available *LeafSnap*, *Flavia* and *Foliage* datasets.

Results: Evaluating the recognition accuracies of *LeafNet* on the *LeafSnap*, *Flavia* and *Foliage* datasets reveals a better performance of *LeafNet* compared to hand-crafted customized systems.

Conclusions: Given the overall species diversity of plants, the goal of a complete automatisation of visual plant species identification is unlikely to be met solely by continually gathering assemblies of customized, specialized and hand-crafted (and therefore expensive) identification systems. Deep Learning CNN approaches offer a self-learning state-of-the-art alternative that allows adaption to different taxa just by presenting new training data instead of developing new software systems.

1. Introduction

The process of plant identification is an important component of typical workflows in plant ecological research. Speeding this task up and making it accessible for non-experts would be highly beneficial, even more so in view of the many threats to and general irreversible decline of plant biodiversity, which can only be addressed by continued concerted efforts in plant biodiversity research and conservation biology. Even for researchers with expert knowledge on given plant taxa, the process of manual plant species identification may be difficult to scale to high-throughput demands, while for non-experts, it may be prohibitively time consuming and error-prone. Given ubiquitous resource limitations (time, money, experts), it may turn out as a bottleneck in many projects.

When faced with an unknown plant, manual identification typically involves navigating a key that consists of a series of prescribed identification steps (usually printed single-access keys, rarely more modern digital interactive keys, see e.g. Stevenson et al., 2003 or Farnsworth et al., 2013). At each identification step, a question about some of the plant's characters needs to be answered. The selected answer out of two (dichotomous key) or more (polytomous key) selectable answers determines the next identification step. Unfortunately, choosing the appropriate answer may not be trivial, when either (i) the key is of suboptimal design (ambiguous questions, contrasting of overlapping ranges, etc.), (ii) the botanical sample lacks relevant characters, (iii) expert knowledge is necessary but not available in order to correctly answer the question. In consequence, the result from a manual identification may suffer from being less accurate or less reproducible. (See e.g. MacLeod et al., 2010 and references therein for a survey of human error in taxon identification.)

In recent years, DNA barcoding has gained momentum and started to unnecessitate some of the traditional identification needs. However, barcoding initiatives are still lacking or reference databases still under construction in most countries. Even more recently, researchers started to more systematically address the issue by providing plant identification tools that employ image recognition technologies (e.g. Cope et al.,

* Corresponding author.
E-mail addresses: barre@cs.uni-bonn.de (P. Barré), stoever@bioinfweb.info (B.C. Stöver), kaimueller@uni-muenster.de (K.F. Müller), steinhag@cs.uni-bonn.de (V. Steinhage).

http://dx.doi.org/10.1016/j.ecoinf.2017.05.005
Received 22 December 2016; Received in revised form 12 May 2017; Accepted 21 May 2017
Available online 22 May 2017

2012; Goëau et al., 2016; Joly et al., 2014; Kadir et al., 2011). In contrast to DNA Barcoding techniques that require DNA extractions from harvested tissue, using computer vision methods for high-throughput identification purposes generally can be non-destructive or less invasive. Also, such methods hold the potential to be easier applicable to herbarium specimens where DNA quality has degraded and, thus, DNA Barcoding is no option.

1.1. Related work

Variations on leaf characteristics are preferably employed in automated plant identification systems using computer vision methods because of leaves being easier observable, accessible and describable compared to other plant organs. Kadir et al., 2011 as well as Cope et al., 2012 and Ahmed et al., 2016, give comprehensive surveys on methods for automated plant identification. However, plant identification is still considered to be a challenging and unsolved problem since all classical computer vision employ hand-crafted methods that are dependent on chosen features given in a natural given extreme diversity of botanical data. For example, Jin et al., 2015, employ a classical image processing chain of image binarization to separate background and the leaf, detection of contours and contour corners, and geometrical derivations of leaf tooth features. The approach was evaluated on eight species and Jin et al., 2015 reported in species specific identification accuracies between 72.8% and 79.3%. But obviously, this approach solely cannot deal with species showing no significant appearances of leaf teeth.

To overcome the naturally given limitations of such model-based approaches, model-free approaches and machine learning methods have been introduced. Wilf et al., 2016, employ the model-free Scale-invariant feature transform (SIFT) approach of Lowe, 2004, to detect and describe in a training set of leaf images the visual appearances of so-called interest points which are assembled in a so-called codebook. A standard approach to supervised learning (i.e., the Support Vector Machine (SVM)) is then trained on the coefficients for all codebook elements and the associated taxonomic labels (families or orders). The learned classification function is then applied to predict familial and ordinal identifications of novel images. The SIFT approach detects and describes interest points based on the appearances of significant local image gradients, i.e., contours, corners, lines, etc. and thus does not model certain features explicitly (model-free approach). Wilf et al., 2016, apply their approach on the vein features of the leaf images and report a classification accuracy of 72.14% for 19 families. However, the images are taken from cleared specimens that are prepared in a laborious way and there is still a limitation to a restricted class of features.

Additionally, machine learning approaches to active learning are employed to meet the challenges of data acquisition, data labelling (i.e., taxonomic identification of training data), and availability of experts. Methods of active learning implement the concept of the user-in-the-loop (UIL) to improve system performance by engaging the human users (or some other information source). In this context, two prominent approaches to mobile participatory sensing and citizen science working on social image data are presented by Kumar et al., 2012, and by Joly et al., 2014 and both use a classical computer vision pipeline.

Kumar et al., 2012, propose a mobile app, called *LeafSnap*, to enable users to identify trees from photographs of their leaves. *LeafSnap* achieves a top-1 recognition rate of about 73% and a top-5 recognition rate of 96.8% for 184 tree species. Joly et al., 2014, propose *Pl@ntNet* as a citizen science approach to speed up the collection and integration of observed botanical image data. The web interface of *Pl@ntNet* offers to employ images of different plant organs (i.e., leaves, fruits, barks, flowers) and of the complete plant to identify a plant. Joly et al., 2014, evaluated *Pl@ntNet* on about half of the plant species of France (2200 species), showing top-5 identification rates of up to 69% for single images. Again, *LeafSnap* and *Pl@ntNet*, are designed with dependencies on the chosen sets of hand-crafted features that had been selected a priori to measure similarities between novel plant images and plant organ images and stored images of known species.

The competition within the ImageCLEF initiative[1] was based on the *Pl@ntView* dataset which focuses on 250 herb and tree species from France area and contains 26,077 pictures showing plant organs and the entire plants Goëau et al., 2013. The task was more related to a retrieval task instead of a pure classification task in order to consider a ranked list of retrieved species rather than a single determination. Yanikoglu et al., 2014 describe the approach of their team from Sabanci Univ. and Okan Univ. (both from Istanbul, Turkey) that had been the winner for images of the so-called "*SheetAsBackground*" category with an average score of 0.607.

The last step towards model-free approaches to plant identification is to get rid of hand-crafted features. In the last years, deep learning convolutional neural networks (CNNs) have seen a significant breakthrough in computer vision, especially in the field of visual object categorization Krizhevsky et al., 2012, due to the rise of efficient general-purpose computing on graphics processing units (GPGPU provides high degrees of parallelization) and the availability of large-scale image data (in publicly available datasets, in the internet, in social media, (specialized) social networks, etc.) that provide the data amount necessary for training deep CNNs with thousands of parameters. An essential advantage of deep CNNs is the automatic learning of task-specific representations of the input data which replace traditional feature-based representations using hand-crafted features.

The deployment of deep CNNs has especially led to a breakthrough in fine-grained visual categorization (FGVC). Fine-grained or so-called subordinate categories are often challenging due to high intra-class variance and low inter-class variance since some categories only differ in detail that only experts notice.

The categorization of plant organs definitely belongs to the FGVC field, since plant organs can show very high variance of visual appearance within the same genus or the same species as well as high similarities across different genera or species.

Lee et al., 2015 presented a CNN approach to taxon identification based on leaf images and reported an average accuracy of 99.7% on a dataset covering 44 species. Zhang et al., 2015 used CNN to classify the Flavia dataset with and obtain an accuracy of 94,69%. Goëau et al., 2016 report on the plant identification task PlantCLEF 2016 that was organized within the ImageCLEF initiative[2] dedicated to the system-oriented evaluation of visual based plant identification. The competition based on the PlantCLEF 2015 dataset[3] which consists out of 113,205 pictures depicting plant organs and entire plants covers 1000 woody and herbaceous species from France and neighbouring countries. Only 8 of 94 research groups succeeded in submitting runs and all employed CNNs. The winning team of the KDE lab of Toyohashi Univ. of Technology, Japan, reached a classification mean average precision of 74.2% (cf. Hang et al., 2016). Aside from academic prototypes and competitions CNN-based approaches to visual categorization and especially to fine-grained visual categorization are on the way to become practical computer vision systems employable by practitioners as depicted by Lu et al., 2016 who apply a CNN-based approach to maize cultivar identification in agriculture.

1.2. Contributions

We present an approach to automated plant identification based on scans and smartphone pictures taken from the leaves of the plants. Our approach employs the technology of convolutional neural networks and

[1] http://www.imageclef.org/(retrieved 31 Aug. 2016)
[2] http://www.imageclef.org/(retrieved 31 Aug. 2016)
[3] http://www.imageclef.org/lifeclef/2015/plant, (retrieved 31 Aug. 2016)

show superior identification results compared with state-of-the-art systems.

2. Materials and methods

We will first present the datasets that we employ for training and evaluation of our deep CNN approach to plant identification using leaf images. We will then proceed to explain the design of our deep learning CNN that we call *LeafNet*.

2.1. Data material and data augmentation

For training and testing the CNNs, we use the *LeafSnap* dataset[4] (Kumar et al., 2012) that shows two subsets: (1) 23 147 so-called lab images (Fig. 1, left), i.e., high-quality images taken of pressed leaves, from the Smithsonian collection with controlled illumination, (2) 7719 so-called field images (Fig. 1, right) taken by mobile devices (iPhones mostly) from different users in outdoor environments. Field images contain varying amounts of blur, noise, illumination patterns, shadows, etc. - but the image taking instructions of the *LeafSnap* project assure that each field image shows a uniform background (provided by holding a sheet of paper behind each leaf while taking the picture). The lab images cover all 185 tree species from the North-eastern United States while the field images cover 184 of the 185 tree species.

The most common approach to reduce overfitting on image data is to enlarge the training dataset synthetically by applying label-preserving transformations to the image data, i.e. position shifting, scaling and rotation of the objects in the image plane as well as brightness and contrast of the images. We apply data augmentation to 21 044 out of the 23 147 *lab images* and to 5580 of the 7719 *field images* enlarging the dataset for training up to 270 161 images.

Since CNNs ask for fixed-sized input images, all RGB images of the *LeafSnap* dataset are standardized in a fully automated manner by downscaling of all images to the fixed size of 256 × 256 pixels. Additionally, the mean of every RGB channel is subtracted from the RGB values of all pixels to centre the dataset to zero and achieve a faster learning of the CNN.

2.2. Deep learning framework and LeafNet system

For *LeafNet* we utilize the deep learning framework *Caffe* (Jia et al., 2014) that is developed by the Berkeley Vision and Learning Center (BVLC) and by community contributors.[5] *Caffe* supports general-purpose computing on graphics processing units (GPGPU) with *NVIDIA* GPUs using the parallel computing platform and application programming interface CUDA[6] with the deep neural network library CuDNN.[7] For all experiments, we employed the NVIDIA GTX 960 graphics card with 4 GB memory and 1024 kernels. The complete system was developed and evaluated with Ubuntu 15.10 (AMD64).

CNNs are special artificial neural networks showing connectivity patterns of their neurons that are inspired by the animal visual cortex. Therefore, CNNs are applied to visual recognition tasks, i.e., the detection, localization and classification of objects in images and video data. The core building blocks of CNNs are the so-called *convolution layers*. Each neuron of a convolution layer responds to stimuli in a restricted region known as its *receptive field*. Its size is called *filter shape* and shows typical values in the range of 3 × 3 to 15 × 15. In the case of a 3 × 3 filter shape each neuron responds to the stimuli of a 3 × 3 portion of the input image. Each convolution layer generates multiple *feature maps* that are new feature-based representations of the original image. Convolution layers are stacked yielding deep CNNs comprising the detection of richer arsenals of usable features and therefore improved recognition rates. While lower convolution layers (the ones close to the input) generate elementary features like edges of different orientations, upper convolution layers derive more complex and task-specific features like special subsampled object regions (Fig. 2).

In contrast to classical computer vision systems that employ hand-crafted algorithms to extract some pre-defined sets of features, CNNs learn automatically to extract features sets that are appropriate for the given visual task. Learning is done by presenting CNNs repeatedly training data, i.e., image data on the one hand and the corresponding solutions (e.g., object classes present in the image data) on the other hand. While training CNNs adapt their internal parameters to learn how to process the visual data to fulfil the given recognition task.

The overall CNN architecture of *LeafNet* utilizes the concept of dimension reduction modules (cf. Simonyan and Zisserman, 2014). In *LeafNet* each module consists of 2 convolutional layers followed by a MAX-pooling layer with filter shape 2 × 2 which halves the feature maps width and height. MAX-pooling layers are essential to allow for deeper but still computational feasible CNNs since they reduce dramatically the number of connections between the last convolution layer and the first full-connected layer.

The overall CNN architecture of *LeafNet* starts with five of these modules followed by a last convolution layer with a MAX-pooling layer of filter shape 2 × 2 and terminates with three fully-connected layers. While the convolutional modules learn (in the training phase) and later perform (in the identification phase) feature extraction, the fully-connected layers learn and later perform the classification, i.e., the derivation of the detected object classes.

Fig. 3 depicts the overall design. The input layer's size is 256 × 256 × 3 to capture the three RGB channels of leaf images of size 256 × 256. The two convolution layers of the first module have filter shape 9 × 9 and those of the second module have filter shape 5 × 5. The convolution layers of all deeper modules have filter shape 3 × 3. After each module, feature maps sizes are reduced by half while the number of features maps is doubled. A last convolutional layer of size 8 × 8 with 768 feature maps is reduced to feature maps of size 4 × 4 by a last max pooling with filter shape 2 × 2. This last convolutional layer is followed by three fully-connected layers where the first two layers consists of 2048 neurons and the last output layer shows 185 neurons, i.e., the number of species to be identified. For the last output layer the well-known softmax function is used to map the resulting c output values for each possible class to probabilities and allow therefore more differentiated rankings of results using the *top-k metric* reporting if the correct class of an input image is within the top-ranked k responses.

All neurons of each convolution layer are spatially tiled over the entire visual field such that neighbouring neurons are located only 1 spatial unit apart. This tiling is said to have stride $S = 1$. Zero-padding is applied to every convolution layer to control the spatial size of the output layers.

Larger filter shapes in the lower layers of the CNN capture larger regions of the feature map and so improve the quality of the convolution. Successive dimension reduction and the growing number of feature maps in the upper layers will provide a growing diversity of more complex and task-specific features as depicted in Fig. 2.

3. Results

We trained the *LeafNet* network on the *LeafSnap* dataset in 200,000 iterations and employed a momentum $\alpha = 0.9$, mini-batches of 10 training examples and started with a learning rate $\eta = 0.001$. After 100,000 iterations the learning rate is decreased by a gamma value $\gamma = 0.1$ resulting in a new value of the learning rate $\eta = 0.0001$ to foster convergence.

Additionally, we trained and evaluated the *LeafNet* CNN on two

[4] http://leafsnap.com/dataset/ (retrieved 31 Aug. 2016)
[5] http://caffe.berkeleyvision.org/(retrieved 31 Aug. 2016)
[6] http://www.nvidia.com/object/cuda_home_new.html (retrieved 31 Aug. 2016)
[7] https://developer.nvidia.com/cudnn (retrieved 31 Aug. 2016)

Fig. 1. A lab image (left) and a field image (right) of the LeafSnap online-dataset (Kumar et al., 2012).

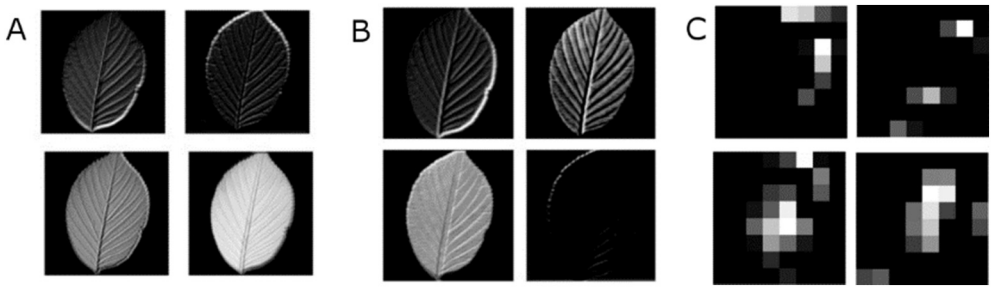

Fig. 2. Some of the feature maps derived by processing a leaf image with a trained CNN. While the first (A) and second (B) convolution layers learn to extract edges and leaf venation, deeper convolution layers (C) derive higher degrees of feature abstraction, i.e. more complex features of interest.

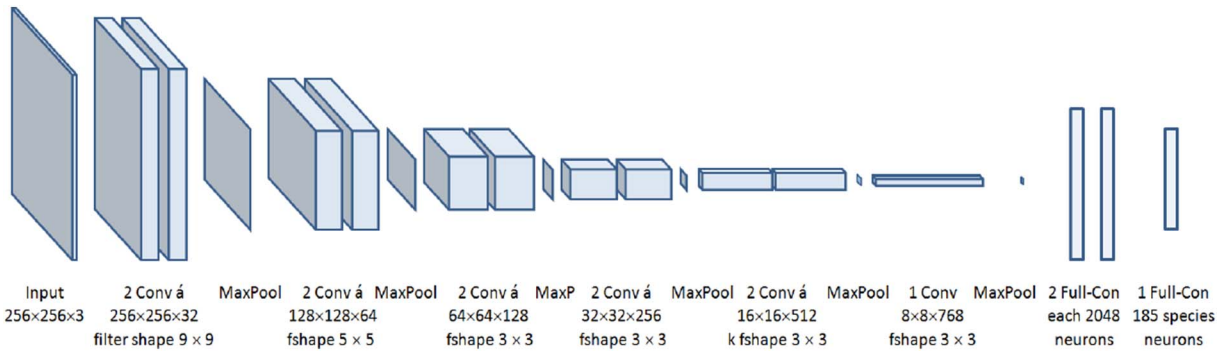

Fig. 3. LeafNet architecture in a simplified visualization (cf. text).

Table 1
Different settings for training of the LeafSnap, the Foliage and the Flavia datasets.

Dataset	No. of images for training before data-augmentation	No. of images for training after data-augmentation	Overall no. of iterations	No. of iterations with decreased η	No. of tested images	No. of tested species
LeafSnap	26,624	270,161	200,000	100,000	2139	184
Foliage	6.000	174,000	100,000	50,000	1200	60
Flavia	1526	44,242	30,000	15,000	381	32

other publicly available state-of-the-art datasets of leaf images, namely the *Foliage* dataset (cf. Kadir et al., 2011) and the *Flavia* dataset[8] (cf. Wu et al., 2007).

Due to different sizes of the datasets, we adapted the training's settings for each dataset as depicted in Table 1.

3.1. Results on the LEAFSNAP dataset

In case of the *LeafSnap* dataset, we only used field images (following Kumar et al., 2012, i.e., photographs taken in the field in the plant's natural environment, rather than in the laboratory) to evaluate our *LeafNet* CNN (cf. Section 2.1). Field images allow to analyze the performance of the system under conditions close to those of a typical user, e.g., the classification of a plant leaf photo taken in the field.

We achieved an average top-1 accuracy of 86.3% and an average top-5 accuracy of 97.8%. The bright diagonal of confusion matrix of the average top-1 accuracy (left part of Fig. 4) also indicates that by far most of the species are identified very well.

Outliers can mostly be explained as species that show strong visual similarities (cf. right part of Fig. 4) or species that are represented in the *LeafSnap* dataset by only few images (making training of such species

[8] [http://flavia.sourceforge.net/(retrieved 31 Aug. 2016)].

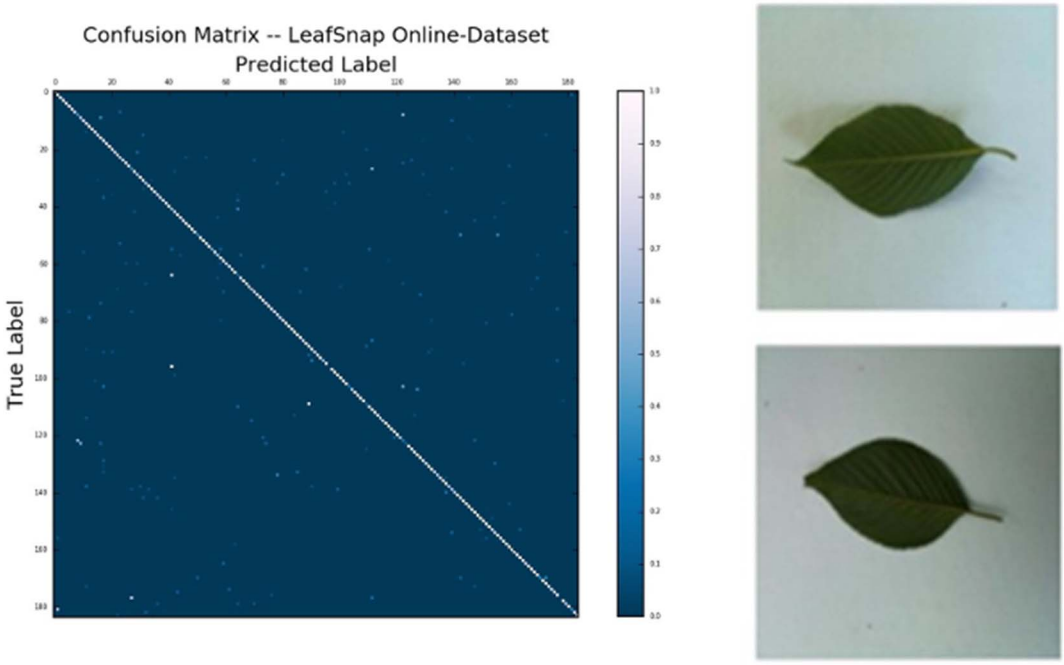

Fig. 4. Left: The confusion matrix for the LeafSnap online dataset. Colours indicate the proportion of leaf images from the actual species that our LeafNet CNN classified into the predicted categories. Right: leaf images of the species *Prunus subhirtella* (top) and *Prunus virginiana* (bottom) (Kumar et al., 2012). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

difficult). For example, Fig. 5 shows the classification for each species of the *LeafSnap* Dataset under the genus *Prunus*. As it shows, the misclassification in other genus (red bin, Other genera) doesn't occurs often, but it appears that most of the misclassified images are classified in another. In a future Work it will be interesting to investigate a way to subclassify species under a high fine-grained condition after a genus classification.

3.2. Results on the FOLIAGE dataset and the FLAVIA dataset

The *Foliage* dataset consists of 60 species, each showing 120 images. Our *LeafNet* CNN achieved an average top-1 accuracy of 95.8% and an average top-5 accuracy of 99.6%. The *Flavia* dataset consists of 32 species and the number of image per species varies from 50 to 60. Our *LeafNet* CNN achieved an average top-1 accuracy of 97.9% and an average top-5 accuracy of 99.9%. The confusion matrices of the average top-1 accuracies of both evaluations (Fig. 6) show again bright diagonals confirming the high average accuracies.

3.3. Overall results and evaluation metrics

Results of automated approaches to taxon identification are reported using different evaluation metrics. To compare with results of other approaches, we explain the most common evaluation metrics and apply these to our overall results.

Average top-n accuracies are commonly used mostly with $n = 1$, 5 or 10. The average top-1 accuracy corresponds to the average recognition accuracy. Given m evaluation samples and y_i as the correct species class for input sample $\mathbf{x}_i$, $i = 1,\cdots,m$, and $j = 1,\cdots n$, the highest ranked species predictions according to their probability values, the precise derivation is:

$$\text{average top-n accuracy} = \frac{1}{m}\cdot\sum_{i=1}^{m}\sum_{j=1}^{n} I\left(f\left(\boldsymbol{x}_i, \boldsymbol{W}\right) = y_i\right),$$

where $I(.)$ is the indicator function returning 1 for true statements and 0 otherwise.

Average scores were used to report the results of the ImageCLEF 2013 plant identification task (Goëau et al., 2013). Given m evaluation

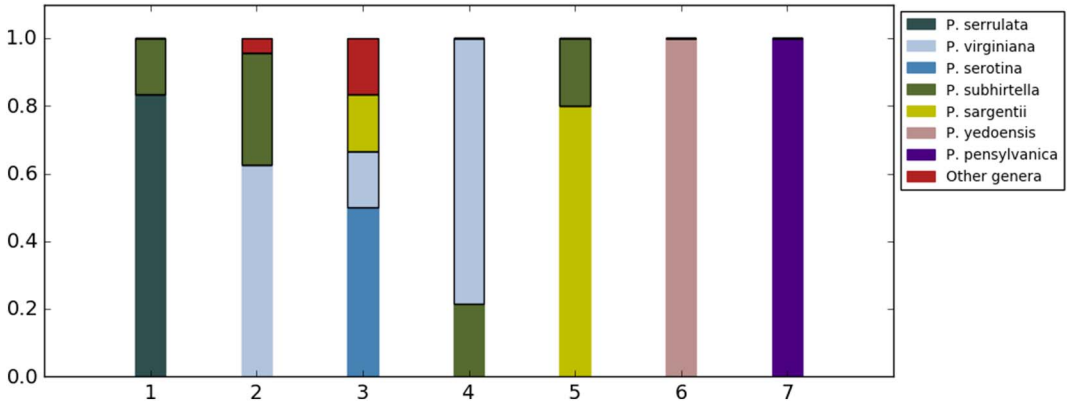

Fig. 5. Distribution of the classification innerhalb within the genus "*Prunus*". The seven *Prunus* species (*P. serrulata* [1], *P. virginiana* [2], *P. serotina* [3], *P. subhirtella* [4], *P. sargentii* [5], *P. yedoensis* [6] and *P. pensylvanica* [7]) have mainly misclassif misclassification within their genus. Only *P. virginiana* [2] und *P. serotina* [3] are also assigned to another genus. (For interpretation of the references to color in this figure the reader is referred to the web version of this article.)

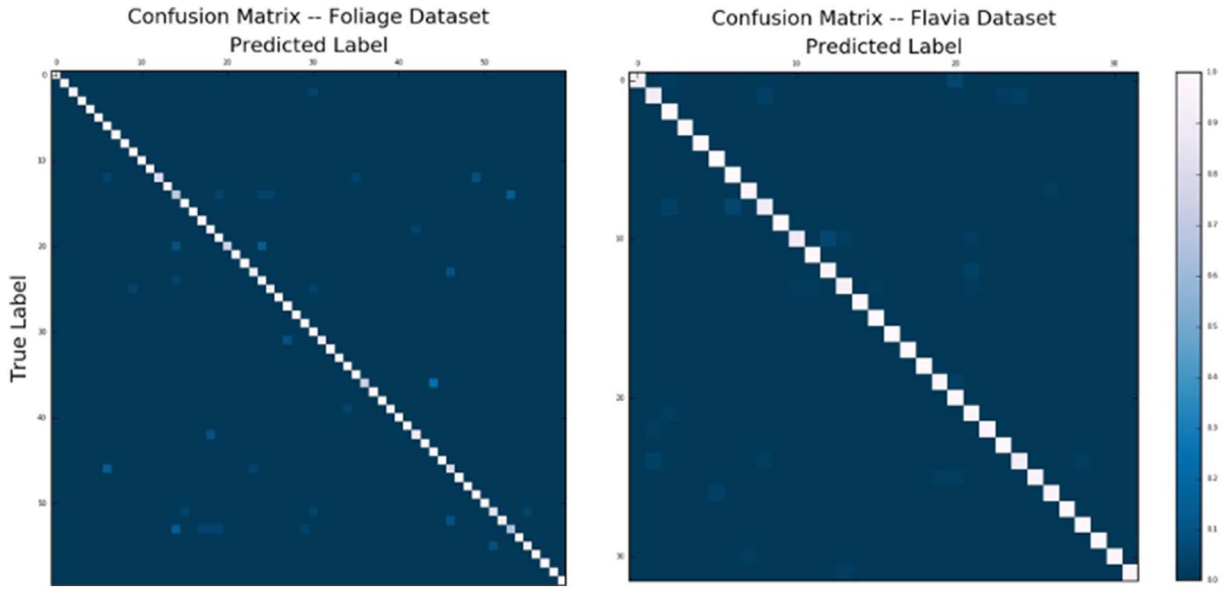

Fig. 6. The confusion matrices for the Foliage and Flavia datasets.

samples and $rank_i$ as the rank of the correct species for sample i, $i = 1,\cdots,m$, the scores represent the mean reciprocal rank (MRR) of the correct species:

$$MRR = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{rank_i}$$

Mean average precision (MAP) has been used to report the results of the ImageCLEF 2016 plant identification task (Goëau et al., 2016). The average precision *AvgPr* (c) is derived for every class c by ranking all the true positives *and* false positives by decreasing probability p ∈ [0;1] and summing up and averaging the precision values of only the true positive predictions:

$$AvgPr(c) = \frac{1}{\#TP}\cdot\sum_{k=1}^{\#P} Pr(k)\cdot I(f(\boldsymbol{x_i}, \boldsymbol{W}) = y_i),$$

where $I(.)$ is the indicator function equaling 1 if the item at rank is a true positive prediction (i.e., the prediction $f(\boldsymbol{x_i}, \boldsymbol{W})$ equals the known classification result y_i), zero otherwise and $Pr(k)$ is the precision at cut-*off* k in the ranking list, i.e., the number of true positives counted so far divided by the number of all positives counted so far. The overall number of ranked prediction (i.e., all positives) is depicted as $\#P$. This way, the MAP are highly degraded by false positive predictions. Then an average precision is derived by averaging the average precision values of all classes (#TP). Table 2 summarizes the evaluation of our *LeafNet* CNN according to the mentioned metrics.

4. Discussion

To judge the results achieved by our deep learning *LeafNet* CNN, it is valuable to relate them to the results achieved by the original contributions that have been applied on the *LeafSnap*, *Foliage* and the *Flavia* datasets.

Table 2
Results on the Flavia dataset, Foliage dataset and LeafSnap-online dataset.

Dataset	No. of species	Top-1 accuracy	Top-5 accuracy	MRR score	MAP score
LeafSnap	184	86.3%	97.8%	92.2%	83.7%
Foliage	60	95.8%	99.6%	97.6%	95.3%
Flavia	32	97.9%	99.9%	98.8%	97.2%

The *Flavia* dataset was processed by Wu et al., 2007 using a conventional approach of a hand-crafted feature extraction and a probabilistic neural network for classification. They report a top-1 accuracy of 90.3% on the *Flavia* dataset. Kadir, 2014, employed a classical computer vision pipeline yielding an accuracy rate of 97.19% when using the *Flavia* dataset and 95.00% when using the Foliage dataset. Kumar et al., 2012, with a classical approach within their mobile *LeafSnap* system achieved a top-1 accuracy of about 73% and a top-5 accuracy of 96.8% for 184 tree species of the *LeafSnap* dataset. The winning research team of the *LifeCLEF plant challenge 2016* reached a classification mean average precision of 74.2% (cf. Hang et al., 2016). However, this MAP result is averaged over results obtained on classifying pictures showing plant organs and the entire plants.

Therefore, this MAP result is difficult to relate to our MAP results between 83.7% and 97.2% (cf. Table 2) and only listed to give a rough statement. The top-1 accuracies are also depicted in Table 3. It must be stressed (again) that the approaches of Wu et al., 2007, Kumar et al., 2012 and Kadir, 2014, employ tailored and hand-crafted feature extraction processing while our *LeafNet* CNN is learning the complete feature extraction processing automatically from training data. In order to show the advantages of an end-to-end learning, we applied the approach from Sharif Razavian et al., 2014 and extract features from the last fully-connected layer of the *OverFeat* CNN (Sermanet et al., 2013) and trained an SVM with the obtained vectors. On the two smaller datasets (*Flavia* and *Foliage*), we obtain better results as *LeafNet* (respectively 98.69% and 98.75% accuracy). However, applying this approach to the larger and by far more challenging *LeafSnap* dataset we achieve a poorer result (79.66%). It also seems that the pre-trained features extracted with the *OverFeat* CNN are sufficiently discriminative to classify scant fine-grained datasets like the *Flavia* dataset and the *Foliage* dataset. But for a highly fine-grained dataset, a specially trained and adapted CNN is needed in order to learn high discriminative and

Table 3
Top-1 accuracies achieved on the Flavia, Foliage and LeafSnap-online datasets.

Dataset	No. of species	Top-1 acc. of our approach	Top-1 acc. of Wu et al., 2007	Top-1 acc. of Kadir, 2014	Top-1 acc. of Kumar et al., 2012
LeafSnap	184	86.3%	/	/	73.0%
Foliage	60	95.8%	/	95.0%	/
Flavia	32	97.9%	90.3%	97.2%	/

specific features.

Furthermore, we extracted features of the last-fully-connected layer of *LeafNet*, trained again a SVM with the gained features with a result of 82.3% accuracy. This approach does not perform better than *LeafNet* (86.3% accuracy). This demonstrates the advantages of a fully end-to-end learning for the given problem of fine-grained classification.

The superior identification results of our end-to-end learned *LeafNet* go also along with more domain specific feature maps, i.e., the derivation of features showing botanical entities like leaf venations and leaf edges. In this context, deconvolution networks offer an opportunity to visualize these automatically derived and domain specific features and offer botanists the opportunity for quality control and domain-specific interpretation. First encouraging results are presented for leaves by Lee et al., 2015. Overall, the results of our *LeafNet* CNN demonstrate that learning features through CNN can provide better feature representations for leaf images compared to hand-crafted features with respect to species identification of plants based on leaf images.

Furthermore, we demonstrate that training of such a CNN with big datasets is possible even with a low-cost hardware solution consisting of a NVIDIA GTX 960 graphics card (~200 US$) running with Ubuntu 15.10 (AMD64) on a conventional desktop PC. Training of our *LeafNet* CNN using the approximately 270,000 leaf images (after data augmentation) lasted about 32 h. But the training is done offline (i.e., before employing *LeafNet* CNN for species identification). The identification process itself performs species identification on 850 leaf images per millisecond.

Because new taxa can be processed by adapting a given CNN architecture just by presenting new training data instead of designing and developing new software systems, the approach is of interest for museums and other institutions to process extensive scientific collections. While the training of extensive datasets is currently time consuming (but we are talking about big data), it can be done in a fully automated way before being employed. *LeafNet* is available at www.leafnet.pbarre.de und released under the BSD-2 Clause Licence.

Acknowledgements

This study was partially funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG, grant no. STE 806/2-1 to V.S., DFG had no influence on study design).

References

Ahmed, N., Khan, U.G., Asif, S., 2016. An automatic leaf based plant identification system. Forensic Sci. Int. 28, 427–430.

Cope, J.S., Corney, D., Clark, J.Y., Remagnino, P., Wilkin, P., 2012. Plant species identification using digital morphometrics: a review. Expert Syst. Appl. 39, 7562–7573.

Farnsworth, E.J., Chu, M., Kress, W.J., Neill, A.K., Best, J.H., Pickering, J., Stevenson, R.D., Courtney, G.W., VanDyk, J.K., Ellison, A.M., 2013. Next-generation field guides. Bioscience 63, 891–899.

Goëau, H., Joly, A., Bonnet, P., Bakic, V., Barthélémy, D., Boujemaa, N., Molino, J.-F., 2013. The imageCLEF plant identification task 2013. In: Presented at the Proceedings of the 2nd ACM International Workshop on Multimedia Analysis for Ecological Data, ACM, pp. 23–28.

Goëau, H., Bonnet, P., Joly, A., 2016. Plant identification in an open-world (LifeCLEF 2016). In: Presented at the CLEF 2016-Conference and Labs of the Evaluation Forum, pp. 428–439.

Hang, S.T., Tatsuma, A., Aono, M., 2016. Bluefield (KDE TUT) at LifeCLEF 2016 Plant Identification Task. Work. Notes CLEF 2016 Conf.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T., 2014. Caffe: Convolutional Architecture for Fast Feature Embedding. ArXiv Prepr. ArXiv14085093.

Jin, T., Hou, X., Li, P., Zhou, F., 2015. A novel method of automatic plant species identification using sparse representation of leaf tooth features. PLoS One 10, e0139482. http://dx.doi.org/10.1371/journal.pone.0139482.

Joly, A., Goëau, H., Bonnet, P., Bakić, V., Barbe, J., Selmi, S., Yahiaoui, I., Carré, J., Mouysset, E., Molino, J.-F., Boujemaa, N., Barthélémy, D., 2014. Interactive plant identification based on social image data. Eco. Inform. 23, 22–34.

Kadir, A., 2014. A model of plant identification system using GLCM, Lacunarity and Shen features. Res. J. Pharm., Biol. Chem. Sci. 5, 1–10.

Kadir, A., Nugroho, L.E., Susanto, A., Santosa, P.I., 2011. Leaf classification using shape, color, and texture features. Int. J. Comput. Trends Technol. 1, 225–230.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks, in: advances in neural information processing systems. Adv. Neural Inf. Proces. Syst. 1097–1105.

Kumar, N., Belhumeur, P.N., Biswas, A., Jacobs, D.W., Kress, W.J., Lopez, I.C., Soares, J.V.B., 2012. Leafsnap: a computer vision system for automatic plant species identification. Comput. Vis.–ECCV 2012, 502–516.

Lee, S.H., Chan, C.S., Wilkin, P., Remagnino, P., 2015. Deep-Plant: Plant Identification with convolutional neural networks. Image process. ICIP 2015. IEEE Int. Conf. IEEE 452–456.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60, 91–110.

Lu, H., Cao, Z., Xiao, Y., Fang, Z., Zhu, Y., 2016. Toward good practices for fine-grained maize cultivar identification with filter-specific convolutional activations. IEEE Trans. Autom. Sci. Eng. 1–33.

MacLeod, N., Benfield, M., Culverhouse, P., 2010. Time to automate identification. Nature 467, 154–155.

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y., 2013. Overfeat: Integrated Recognition, Localization and Detection Using Convolutional networks. ArXiv Prepr. ArXiv13126229.

Sharif Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S., 2014. CNN features off-the-shelf: an astounding baseline for recognition. In: Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 806–813.

Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-scale Image Recognition. ArXiv Prepr. ArXiv14091556.

Stevenson, R.D., Haber, W.A., Morris, R.A., 2003. Electronic field guides and user communities in the eco-informatics revolution. Conserv. Ecol. 7, 3.

Wilf, P., Zhang, S., Chikkerur, S., Little, S.A., Wing, S.L., Serre, T., 2016. Computer vision cracks the leaf code. Proc. Natl. Acad. USA 113, 3305–3310.

Wu, S.G., Bao, F.S., Xu, E.Y., Wang, Y.-X., Chang, Y.-F., Xiang, Q.-L., 2007. A leaf recognition algorithm for plant classification using probabilistic neural network. 2007 IEEE Int. Symp. Signal Process. Inf. Technol. 11–16.

Yanikoglu, B., Aptoula, E., Tirkaz, C., 2014. Automatic plant identification from photographs. Mach. Vis. Appl. 25, 1369–1383.

Zhang, C., Zhou, P., Li, C., Liu, L., 2015. A convolutional neural network for leaves recognition using data augmentation. In: Presented at the Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), 2015 IEEE International Conference on, IEEE, pp. 2143–2150.