Pattern Recognition 71 (2017) 1–13

Contents lists available at ScienceDirect

Pattern Recognition

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

How deep learning extracts and learns leaf features for plant classification

Sue Han Lee[a], Chee Seng Chan[a,*], Simon Joseph Mayo[b], Paolo Remagnino[c]

[a] Centre of Image & Signal Processing, Faculty of Computer Science and Information Technology, University of Malaya, Malaysia
[b] Herbarium, Royal Botanic Gardens, TW9 3AE, United Kingdom
[c] Faculty of Science, Engineering and Computing, Kingston University, KT1 2EE, United Kingdom

ARTICLE INFO

Article history:
Received 11 November 2016
Revised 23 April 2017
Accepted 13 May 2017
Available online 18 May 2017

Keywords:
Plant recognition
Deep learning
Feature visualisation

ABSTRACT

Plant identification systems developed by computer vision researchers have helped botanists to recognize and identify unknown plant species more rapidly. Hitherto, numerous studies have focused on procedures or algorithms that maximize the use of leaf databases for plant predictive modeling, but this results in leaf features which are liable to change with different leaf data and feature extraction techniques. In this paper, we learn useful leaf features directly from the raw representations of input data using Convolutional Neural Networks (CNN), and gain intuition of the chosen features based on a Deconvolutional Network (DN) approach. We report somewhat unexpected results: (1) different orders of venation are the best representative features compared to those of outline shape, and (2) we observe multi-level representation in leaf data, demonstrating the hierarchical transformation of features from lower-level to higher-level abstraction, corresponding to species classes. We show that these findings fit with the hierarchical botanical definitions of leaf characters. Through these findings, we gained insights into the design of new hybrid feature extraction models which are able to further improve the discriminative power of plant classification systems. The source code and models are available at: https://github.com/cs-chan/Deep-Plant.

1. Introduction

Computational botany consists of applying innovative computational methods to help progress on an age-old problem, i.e. the identification of the estimated 400,000 species of plants on Earth [1]. This interdisciplinary approach combines botanical data and species concepts with computational solutions for classification of plants or parts thereof and focuses on the design of novel recognition methods. These are modelled using botanical data, but are extendable to other large repositories and application domains. Plant species identification is a subject of great importance in many fields of human endeavour, including such areas as agronomy, conservation, environmental impact, natural product and drug discovery and other applied areas [2,3].

Advances in science and technology now make it possible for computer vision approaches to assist botanists in plant identification tasks. A number of approaches have been proposed in the literature for automatic analysis of botanical organs, such as leaves and flowers [4–6]. In botany, leaves are almost always used to supply important diagnostic characters for plant classification and in some groups exclusively so. Since the early days of botanical science, plant identification has been carried out with traditional text-based taxonomic keys that use leaf characters, among others. For this reason, researchers in computer vision have used leaves as a comparative tool to classify plants [7–10]. Characters such as shape [11–13], texture [14–16] and venation [17,18] are the features most generally used to distinguish the leaves of different species. The history of plant identification methods, however shows that existing plant identification solutions are highly dependent on the ability of experts to encode domain knowledge. For many morphological features pre-defined by botanists, researchers use hand-engineering approaches for their characterization. They look for procedures or algorithms that can get the most out of the data for predictive modeling. Then, based on their performance, they justify the subset of features that are most important to describe leaf data. However, these features are liable to change with different leaf data or feature extraction techniques. This observation therefore raises a few questions: (1) *In general, what is the best subset of features to represent leaf samples for species identification?* (2) *Can*

* Corresponding author.
E-mail addresses: leesuehan@siswa.um.edu.my (S.H. Lee), cs.chan@um.edu.my (C.S. Chan), s.mayo@kew.org (S.J. Mayo), p.remagnino@kingston.ac.uk (P. Remagnino).

http://dx.doi.org/10.1016/j.patcog.2017.05.015
0031-3203/

we quantify the features needed to represent leaf data? We want to answer these questions in order to solve the ambiguity surrounding the subset of features that best represent leaf data.

In the present study, we propose the use of deep learning (DL) for reverse engineering of leaf features. We first employ one of the DL techniques – Convolutional Neural Networks (CNN) to learn a robust representation for images of leaves. Then, we go deeper into exploring, analyzing, and understanding the most important subset of features through feature visualization techniques. We show that our findings convey an important message about the extent and variety of the features that are particularly useful and important in modeling leaf data.

In this paper, we present several major contributions:

1. We define a way to quantify the features necessary to represent leaf data (Section 4). We first train a CNN based on raw leaf data, then use a Deconvolutional Network (DN) approach to find out how the CNN characterizes the leaf data.
2. We experimentally show that shape is not a dominant feature for leaf representation but rather the different orders of venation (Section 4.3).
3. We quantify the characteristics of features in each CNN layer and find that the network exhibits layer-by-layer transition from general to specific types of leaf feature. We find that this effect emulates the botanists' character definitions used for plant species classification (Section 5).
4. We show that CNNs trained on whole leaves and leaf patches exhibit different contextual information of leaf features. We categorise them into global features that describe the whole leaf structure and local features that focus on venation (Sections 4.3 and 5).
5. We propose new hybrid global-local feature extraction models for leaf data, which integrate information from two CNNs trained using different data formats extracted from the same species (Section 6).
6. We demonstrate that our proposed hybrid global-local feature extraction models can further boost the discriminative power of plant classification systems (Section 6.2.1).

Our paper begins with an introduction to deep learning. Next, we proceed to a critical and comprehensive review of existing methods and a description of the context of plant identification - i.e. how species are delimited by botanists using morphology. Then, we introduce the idea of deep learning for automatic processing and classification in order to learn and discover useful features for leaf data. We describe how computational methods can be adapted and learnt using visual attention. The universal occurrence of variability in natural object kinds, including species, will be described, showing first how it can confound the classification task, but also how it can be exploited to provide better solutions by using deep learning.

2. Deep learning

Deep learning is a class of techniques in machine learning technology, consisting of multiple processing layers that allow representation learning of multiple level data abstraction. The gist of DL is its capacity to create and extrapolate new features from raw representations of input data without having to be told explicitly which features to use and how to extract them.

In the plant identification domain, numerous studies have focused on procedures or algorithms that maximize the use of leaf databases, and this always leads to a norm that leaf features are liable to change with different leaf data and feature extraction techniques. Heretofore, we have been engaged with ambiguity surrounding the subset of features that best represent the leaf data. Hence, in the present study, instead of delving into the creation of feature representation as in previous approaches, we reverse engineer the process by asking DL to interpret and elicit the particular features that best represent the leaf data. By means of these interpretation results, we are able to perceive the cognitive complexities of vision for leaves as such, reflecting the trivial knowledge researchers intuitively deploy in their imaginative vision from the outset.

3. Related studies

In this section, we describe various feature extraction methods that have been proposed to classify species based on different leaf features.

Shape. Most studies use shape recognition techniques to model and represent the contour shape of the leaf. In one of the earliest papers, Neto et al. [11] introduced Elliptic Fourier and discriminant analyses to distinguish different plant species based on their leaf shape. Next, two shape modeling approaches based on the invariant-moments and centroid-radii models were proposed [19]. Du et al. [20] proposed combining geometrical and invariant moments features to extract morphological structures of leaves. Shape Context (SC) and Histogram of Oriented Gradients (HOG) have also been used to attempt to create a leaf shape descriptor [12,13]. Recently, Aakif and Khan [21] proposed using different shape-based features such as morphological characters, Fourier descriptors and a newly designed Shape-Defining Feature (SDF). Although the algorithm showed its effectiveness in baseline dataset like Flavia [5], the SDF is highly dependent on the segmented result of leaf images. Hall et al. [8] proposed using Hand-Crafted Shape (HCS) and Histogram of Curvature over Scale (HoCS) [7] to analyse leaves. Zhao et al. [22] proposed a new counting-based shape descriptor, namely independent-IDSC(I-IDSC) features, to recognize simple and compound leaves. Apart from studying the whole shape contour of the leaf, some studies [9,23] analysed leaf margins for species classification. There are also some groups of researchers who are incorporating plant identification into mobile computing technology such as *Leafsnap* [7] and *Apleafis* [24].

Texture. Texture is another major field of study in plant identification. It is used to describe the surface of the leaf based on the pixel distribution over a region. One of the earliest studies [25] applied multi-scale fractal dimension to plant classification. Next, Cope et al. [16] proposed using Gabor co-occurrences in plant texture classification. Rashad et al. [26] employed a combined classifier – Learning Vector Quantization (LVQ) together with the Radial Basis Function (RBF) – to classify and recognize plants based on textural features. Olsen et al. [27] proposed using rotation and a scale invariant HOG feature set to represent regions of texture within leaf images. Naresh and Nagendraswamy [14] modified the conventional Local Binary Patterns (LBP) approach to consider the structural relationship between neighboring pixels, replacing the hard threshold approach of basic LBP. Tang et al. [15] introduced a new texture extraction method, based on the combination of Gray Level Co-Occurrence Matrix (GLCM) and LBP, to classify tea leaves.

Venation. Identification of leaf species from their venation structure is widely used by botanists. In computer vision, Charters et al. [17] designed a novel descriptor called EAGLE. It comprises five sample patches that are arranged to capture and extract the spatial relationships between local areas of venation. They showed that a combination of EAGLE and SURF was able to boost the discriminative ability of feature representation. Larese et al. [18] recognised legume varieties based on leaf venation. They first segmented the vein pattern using Hit or Miss Transform (UHMT), then used LEAF GUI measures to extract a set of features for veins and areoles. The latest study [30] attempted deep learning in plant identification using vein morphological patterns. They first extracted the vein patterns using UHMT, and then trained a CNN

Table 1
Summary of related studies.

Publications	Year	Method	Features			
			Shape	Texture	Color	Venation
Neto et al. [11]	2006	Elliptic Fourier + Discriminant analyses	✓	–	–	–
Du et al. [20]	2007	Geometrical calculation + Moment invariants	✓	–	–	–
Backes and Bruno [25]	2009	Multi-scale fractal dimension	–	✓	–	–
Cope et al. [16]	2010	Gabor Co-Occurrences	–	✓	–	–
Xiao et al. [13]	2010	HOG + MMC	✓	–	–	–
Beghin et al. [28]	2010	Contour signature + Sobel	✓	✓	–	–
Chaki and Parekh [19]	2011	Moment invariants + Centroid-radii model	✓	–	–	–
Rashad et al. [26]	2011	LVQ + RBF	–	✓	–	–
Mouine et al. [12]	2012	Advanced SC + Hough, Fourier and Edge Oriented Histogram	✓	✓	–	–
Cope and Remagnino [23]	2012	DTW (leaf margin)	✓	–	–	–
Kumar et al. [7]	2012	HoCS	✓	–	–	–
Ma et al. [24]	2013	Wavelet + PHOG	✓	–	–	–
Kadir et al. [10]	2013	Geometrical calculation + Polar Fourier Transform (Shape) + Color moments (Color) + Fractal measure - lacunarity (Texture)	✓	✓	✓	–
Charters et al. [17]	2014	EAGLE	–	–	–	✓
Larese et al. [18]	2014	UHTM + LEAF GUI	–	–	–	✓
Aakif and Khan [21]	2015	Geometrical calculation + Fourier descriptors + SDF	✓	–	–	–
Kalyoncu and Toygar [9]	2015	Margin descriptors + Moment Invariants + Geometrical calculation	✓	–	–	–
Hall et al. [8]	2015	HCS + HoCS	✓	–	–	–
Zhao et al. [22]	2015	I-IDSC	✓	–	–	–
Tang et al. [15]	2015	LBP + GLCM	–	✓	–	–
Olsen et al. [27]	2015	HOG	–	✓	–	–
Chaki et al. [29]	2015	Gabor filter + GLCM + curvelet transform	✓	✓	–	–
Naresh and Nagendraswamy [14]	2016	Modified LBP	–	✓	–	–
Grinblat et al. [30]	2016	UHTM + CNN	–	–	–	✓

to recognise them using a central patch of leaf images. In addition, a considerable amount of research has used combinations of features to represent leaves. For example: attempts to combine shape and texture [28,29] and with the addition of color features [10]. A summary of our literature review is provided in Table 1.

As Table 1 shows, leaf shape features have been chosen and tested in almost 62.5% of plant identification studies, much exceeding the use of other features. This is because they are the easiest and most obvious features for distinguishing species, particularly for non-botanists who have limited knowledge of plant characters. Nevertheless, quite a number of publications used texture features as well (approximately 41.7%) because some species are difficult or impossible to differentiate from one another using only shape due to their similar leaf contours. Although they were shown to be successful, the performance of these approaches is highly dependent on a chosen set of hand-engineered features. In other words, these hand-crafted features are liable to change with different leaf data and feature extraction techniques, which confounds the search for an effective subset of features to represent leaf samples in species recognition studies. With this background, we provide in this paper a solution for the quantification of prominent leaf features.

A preliminary version of this work was presented earlier [31]. The present work adds to the initial version in significant ways. Firstly, we quantify the characteristics of features in each CNN layer and find that the network exhibits layer-by-layer transition from general to specific types of leaf feature. Secondly, we propose new hybrid global-local feature extraction models for leaf data, which integrate information from two CNNs trained using different data formats extracted from the same species. We also extend the original experiments from using MalayaKew to the baseline Flavia Leaf dataset [5].

4. Distinguishing features

In this section, we explain the methodology that we employ to interpret the best subset of leaf features. We first choose one of the DL techniques, namely CNN, to learn a robust representation of leaf images. Later, we show the use of DN to venture into each CNN layer and interpret its neuron computation to quantify the prerequisite features for leaf representation. Fig. 1 depicts the overall framework of our approach.

4.1. Convolutional neural networks

Our CNN model for selecting subsets of leaf features is based on the model proposed in [32] the architecture of which is summarised in Table 2. Rather than training a new CNN architecture, we re-used the pre-trained network because (1) it is widely known that features extracted from the activation of a CNN trained in a fully supervised manner in large-scale object recognition studies can be re-purposed for a novel generic task [33], (2) our training set is not as large as the ILSVRC2012 dataset - as indicated in [34], the performance of the CNN model is highly dependent on the size and level of diversity of the training set, and (3) among the many proposed object classification networks at our disposal, we select the most light-weight and simple network structure to test our concept.

In the convolution layer, feature maps computed in the previous layer are convolved with a set of weights, the so-called filters. That is, the feature map of channel i at layer l, $Y_i^{(l)}$ is computed as: $Y_i^{(l)} = \sum_{j=1}^{m^{(l-1)}} K_{j,i} * Y_j^{(l-1)}$ where K are the filters and $i = 1, 2, \cdots, m^{(l)}$. The resulting feature maps are then passed through a non-linearity unit which is the rectified linear unit (RELU). Next, in the

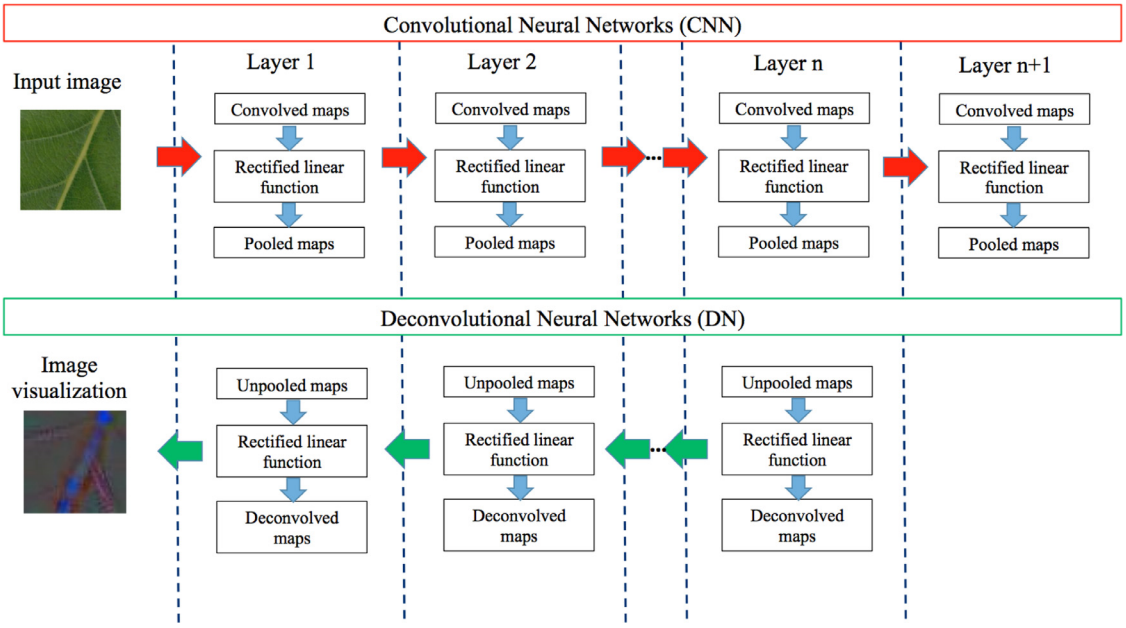

Fig. 1. Our deep learning framework shown in a bottom-up and top-down way to study and understand plant identification. Best viewed in electronic form.

Table 2
CNN architecture used for selection of leaf feature subsets. First, second and third row indicates layer name, number of channels and filter size respectively.

conv1	pool1	conv2	pool2	conv3	conv4	conv5	pool5	fc6	fc7	fc8
96	96	256	256	384	384	256	256	4096	4096	1000
11 × 11	3 × 3	5 × 5	3 × 3	3 × 3	3 × 3	3 × 3	3 × 3	–	–	–

pooling layer, each feature map is subsampled with max pooling over a $q \times q$ contiguous region to produce the so-called pooled maps. After performing convolution and pooling in the fifth layer, the output is then fed into fully-connected layers to perform the classification.

We train our model using *Caffe* [35] framework. For the parameter setting in training, we employ step learning policy. The learning rate was initially set to 10^{-3} for all layers to accept the newly defined last fully connected layer set to 10^{-2}. It is higher than other layers due to the weights being trained starting from random. The learning rate was then decreased by a factor of 10 every 20K iteration and was stopped after 100K iterations. The units of the third fully connected layer (fc8) were changed according to the number of classes of training data. We set the batch size to 50 and momentum to 0.9. We applied L_2 weight decay with penalty multiplier set to 5×10^{-4} and dropout ratio set to 0.5, respectively.

4.2. Deconvolutional network

The CNN model learns and optimises the filters in each layer through the back propagation mechanism. These learned filters extract important features that uniquely represent the input leaf image. Therefore, in order to understand why and how the CNN model operates, filter visualisation is required to observe the transformation of the features, as well as to understand the internal operation and the characteristics of the CNN model. Moreover, we can identify the unique features in the leaf images that are deemed important to characterize a plant by this process.

To quantify the prerequisite features for a leaf image, we attempt to: (1) interpret the function computed by individual neuron/filters, 2) examine the overall function computed in convolution layers composed of multiple neurons. The first attempt is to find out the local response of each filter. It provides us with an intuition concerning the portion of the leaf structure that is important for recognition. Zeiler and Fergus [36] introduced a multi-layered DN that enables us to interpret the function computed by individual neurons by projecting the feature maps back to the input pixel space. Specifically, the feature maps from layer l are alternately deconvolved and unpooled continuously down to the input pixel space. That is, the projected feature map of channel i at layer $l-1$, $Y_i^{(l-1)}$ is computed as: $Y_i^{(l-1)} = \sum_{j=1}^{m^{(l)}} (K_{j,i})^T * Y_j^{(l)}$ where K are the filters and $i = 1, 2, \cdots, m^{(l-1)}$.

Another approach is to examine the overall function computed in a convolution layer composed of multiple neurons. The purpose is to examine areas of overall highest activation across all feature maps for the layer l. Using the reconstructed image, we can observe the highly activated regions of the leaf in that layer. In order to do this, we extend the previous approach [36], proposing a strategy named as **V1**. For all the absolute activations in a layer l, we consider only the first S largest pixel values with the rest set to zero and projected down to pixel space to reconstruct an image defined as: $Y_{i_s}^{(l-1)} = \sum_{j=1}^{m^{(l)}} (K_{j,i})^T * Y_j^{(l)}$ where $S = 1, 2, \cdots, size(Y_j^{(l)})$. With this, we can observe the highly activated regions of the leaf in that layer. Both approaches require a network trained by a leaf dataset and running data through that network for model function interpretation.

4.3. Malayakew dataset

A new leaf dataset, named as the MalayaKew (MK) Leaf Dataset[1] consisting of 44 classes collected at the Royal Botanic Gardens, Kew, England, is employed in the experiment. A dataset

[1] http://web.fsktm.um.edu.my/~cschan/downloads_MKLeaf_dataset.html .

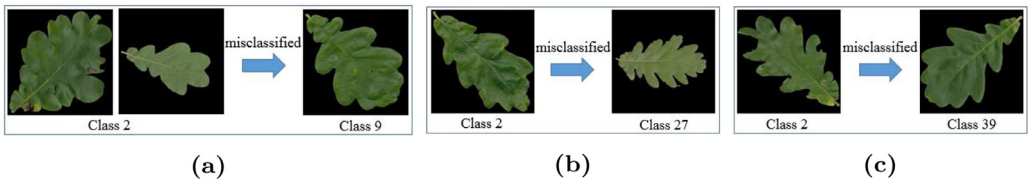

Fig. 2. Failure analysis of the CNN model in D1. Best viewed in electronic form.

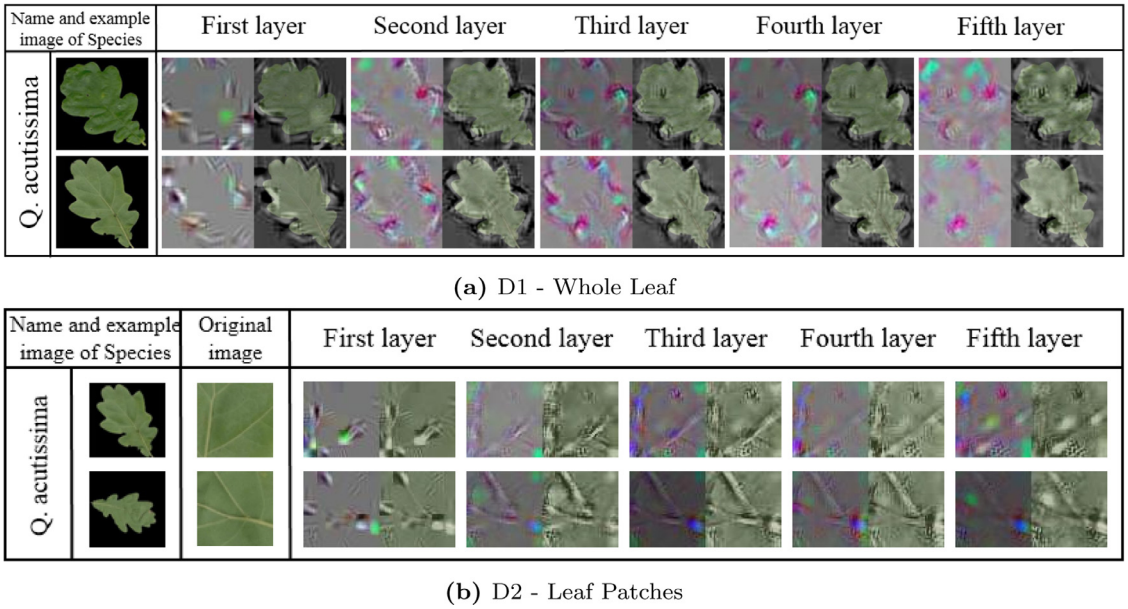

Fig. 3. Feature visualisation using **V1**. This shows that shape (feature) is chosen in D1, while venation and the divergence between different venation orders (feature) are chosen in D2. Best viewed in colour.

Table 3
Performance comparison on the MK leaf dataset with different classifiers. MLP = Multilayer Perceptron, SVM = Support Vector Machine, and RBF = Radial Basis Function.

Feature	Classifier	Acc
From Deep CNN (D1)	**MLP**	**0.977**
From Deep CNN (D1)	**SVM (linear)**	**0.981**
From Deep CNN (D2)	**MLP**	**0.995**
From Deep CNN (D2)	**SVM (linear)**	**0.993**
LeafSnap [7]	SVM (RBF)	0.420
LeafSnap [7]	NN	0.589
HCF [8]	SVM (RBF)	0.716
HCF-ScaleRobust [8]	SVM (RBF)	0.665
Combine [8]	Sum rule (SVM (linear))	0.951
SIFT [37]	SVM (linear)	0.588

(D1) is prepared to compare the performance of the trained CNN. That is, we use leaf images as a whole where in each leaf image, foreground pixels are extracted using the HSV colour space information. To enlarge the D1 dataset, we rotate each leaf image in 7 different orientations, e.g. 45°, 90°, 135°, 180°, 225°, 270° and 315°. We then randomly select 528 leaf images for testing and 2288 images for training. In this experiment, the top-1 classification accuracy is computed to infer the robustness of the system: $Acc = Tr/Tn$ where Tr = number of true species predictions, Tn = total number of images tested.

4.3.1. Results and failure analysis - D1

In this section, we present a comparative performance evaluation of the CNN model for plant identification. From Table 3, it is noticeable that performance of the features learnt from the CNN model (98.1%) is better than state-of-the-art solutions [7,8,37] which employed carefully chosen hand-crafted features, even when different classifiers are used. We performed failure analysis and observed that most of the misclassified leaves are from Class 2(4 misclassified), followed by Class 23(3), Class 9 & 27(2 each), and Class 38(1). From our investigation as illustrated in Fig. 2, the leaves of *Q. robur f. purpurascens* (i.e. Class 2) that were misclassified as *Q. acutissima* (i.e. Class 9) , *Q. rubra Aurea* (i.e. Class 27) and *Q. macranthera* (Class 39), respectively, have almost the same outline shape as those of Class 2. The remaining misclassifications of testing images were also found to have resulted from the same cause.

In order to further understand how and why the CNN fails, we delve into the internal operation and behaviour of the CNN model via V1 strategy. We evaluate the single largest pixel value across the feature maps. Our observation of the reconstructed images in Fig 3a shows that the highly activated parts occur in the shape of the leaves. So, we deduce that leaf shape is not a good choice for identifying plants.

4.3.2. Results and failure analysis - D2

We carried out further investigations by building a variant dataset (D2), where we manually crop each leaf image in the D1 dataset into patches within the area of the leaf (so that leaf shape is excluded). This investigation is two-fold. On the one hand, we wish to determine the precision of the plant identification classifier when the leaf shape is excluded, and on the other, we would like to find out if plant identification could achieve using on a patch of the leaf. Since the original images range from 3000 × 3000 to 500 × 500, three different leaf patch sizes (500 × 500, 400 × 400 and 256 × 256) were chosen. Similarly, we increased the diversity of the leaf patches by rotating them in the same manner as for

D1. We randomly selected 8800 leaf patches for testing and 34,672 patches for training.

In Table 3, we can see that the top-1 accuracy result of the CNN model trained using D2 (99.5%) is higher than that obtained using D1 (97.7%). Again, we perform the visualisation via V1 strategy as depicted in Fig. 3b to understand why the CNN trained with D2 has a better performance. From layer to layer, we notice that the activation part falls not only on the primary venation but also on the secondary venation and the divergence between different orders of venation. Therefore, we can deduce that different orders of venation are more robust features for plant identification. This also agrees with some studies [38,39] which highlight the potential that quantitative leaf venation data have to revolutionize the plant identification task. Existing studies that have employed venation to carry out plant classification are [18,40–43]. However, unlike these solutions, we automatically learned the venation of different orders, while these authors used a set of heuristic rules that are hard to replicate.

We also analysed the drawbacks of the CNN model with D2 and observed that most of the misclassified patches are from Class 9(18 misclassified), followed by Class 2(13), Class 30(5), Class 28(3) and Class 1 , 31 and 42(1 each). The contributing factor to misclassification seems to be the condition of the leaves, where the samples are noticeably affected by environmental factors resulting in wrinkled surfaces and insect damage.

4.4. Discussion

In this experiment, we gain two important intuitions regarding leaf features. Firstly, leaf shape alone is not a good choice for identifying plants because of the common occurrence of similar leaf contours, especially in closely related species. In these situations, venation is a more powerful discriminating feature. In the range of characters used by plant taxonomists, shape and venation are usually used together for characterizing species, and venation is treated hierarchically, with the major veins pattern constituting one character and the minor vein pattern representing another [44]. However, traditional morphological verbal description has limited power to characterize the subtleties of fine venation patterns and in particular its variation.

Secondly, these findings reaffirm the superiority of learned features of leaves based on DL. Our approach discovers more efficient discriminating features than those used for plant identification in previous studies, in which researchers have focused on primarily on shape features because of their convenience. Using DL, we can overcome the inadequacy of shape alone and explore other kinds of characters presented by leaf images.

At this stage, we can see the advantages of CNN in discovering discriminatory features of leaves. However, doubt remains whether it is sufficient to use just venation features for all kind of leaves, and what CNN actually learns in each layer in order to determine the venation features. To clarify these uncertainties, we explore the insights of CNN layers in the following section. We demonstrate how CNN actually works in finding the most distinctive subset of features for leaves and illustrate how it emulates the orthodox basis of descriptive botanical classification used in species distinction.

5. Insights of CNN

In this section, we aim to explore deeper into local response of filters in each convolution layer in order to understand how CNN works in finding the prominent subset of leaf features. This time, we evaluate based on the well-known baseline leaf database – the Flavia dataset [5] in order to show the consistency of CNN performance in different leaf databases. We first quantitatively compare the performance of CNN features with other state-of-the-art methods. Then, we delve into qualitative analysis of the filter response in each convolution layer through the DN approach [36].

Table 4
Performance comparison on the Flavia leaf dataset. FD = Fourier descriptors, SDF = Shape defining features, RF = Random forest, NN = Nearest neighbors and ANN = Artificial neural network.

Feature	Classifier	Average accuracy
From Deep CNN	**MLP**	**0.994**
HCF [8]	RF	0.912
HCF-ScaleRobust [8]	RF	0.898
Combine [8]	Sum rule (RF)	0.973
Morphological,FD,SDF [21]	ANN	0.960
HOG (Multi-scale window) [27]	Gaussian SVM	0.947
Modified LBP [14]	NN	0.976

5.1. Quantitative analysis

In this section, the baseline classification performance for different features proposed was compared using the original set of leaf images from the Flavia dataset. We considered all the leaf samples from each species class, from which 10 samples were selected at random for testing. We repurposed our CNN model to classify 32 classes by altering the last fully connected layer to 32 neurons. Then we optimized the network by fine tuning all the layers end-to-end using the same training algorithm as mentioned in Section 4.1. We input the whole leaf image into our CNN architecture for training as well as testing without cropping them into patches. As a performance metric, we evaluated our system based on the average accuracy that was previously presented in other state-of-the-art methods.

In Table 4, we can see that the classification accuracy of the CNN model achieved the highest average accuracy of 99.4% using MLP classifier, and that it outperforms the state-of-the-art methods either using shape and statistical features [8,21] or texture features [14,27]. Based on these empirical results, we demonstrate that features learned in an unsupervised way, without being imposed by heuristic rules, are more powerful and distinctive for representing leaves that are highly variable in all kinds of leaf characters. Next, we reveal the features learned in each convolutional layer through the DN approach.

5.2. Qualitative analysis

In Section 4, using the V1 strategy on Malayakew dataset, we analysed the global response of filters in each convolution layer. In this section, in order to gain insights into CNN, we further explore the local responses of individual filters in each convolution layer. We randomly subsample some of the feature maps/channels in each layer and reconstruct them back to image pixels to reveal the structures within each patch that stimulated a particular feature map using the DN approach [36]. We also run through all the training samples, and subsequently discover which portions of the training images caused the firing of neurons. By doing this, we can improve our understanding of the transformation of the features learned in each layer and realise the characteristic of each layer in the CNN. Fig. 4 shows the feature visualisation of layer 1. We can see that some of the filters learned are similar to a set of Gabor-like filter banks in different orientations, while some of them depict color areas. This shows that the first layer of the CNN tends to extract low-level features like edges and colors.

Next, we proceed to analyse layers 2–4. In Fig. 5, we show the top two image patches from the training set that caused the highest activations in a random subset of channels in layers 2–4.

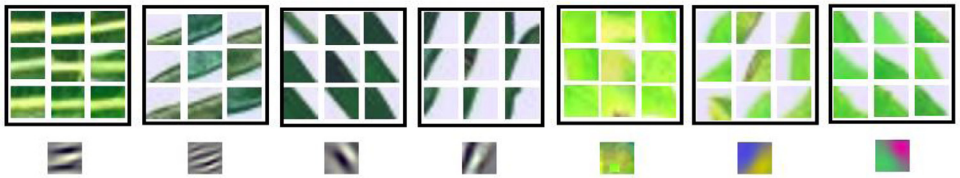

Fig. 4. Feature visualisation of layer 1. The upper row shows the top nine image patches from the training set that caused the highest activations for the selected channels. The lower row shows their deconvolutions.

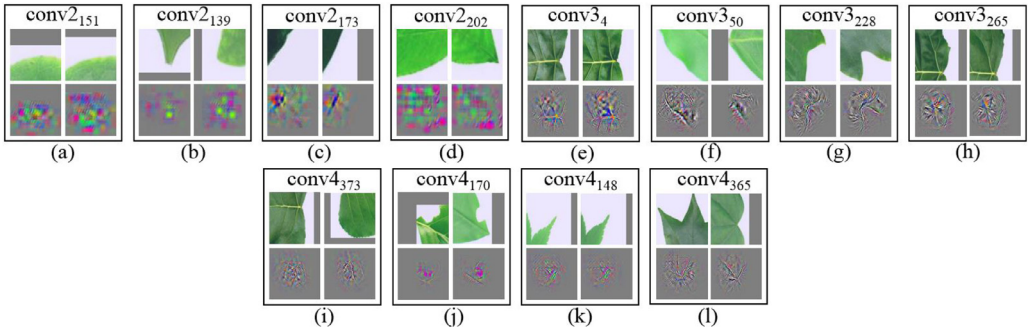

Fig. 5. The top two image patches from the training set that caused the highest activations in a random subset of channels in layers 2–4. Best viewed in electronic form.

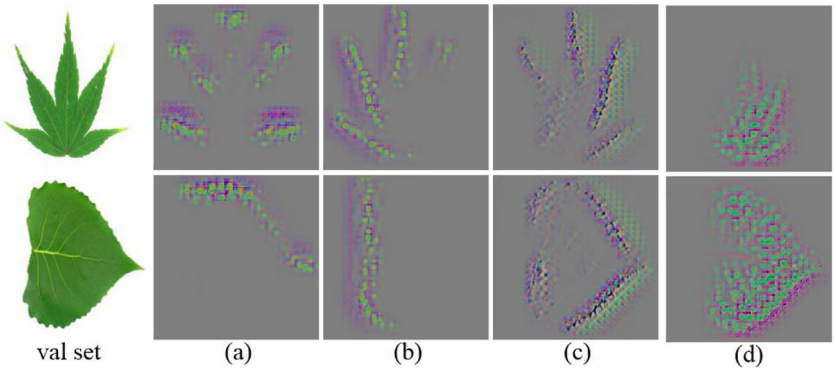

Fig. 6. Each column (a), (b), (c) and (d) depicts deconvolution results of channels $conv2_{151}$, $conv2_{139}$, $conv2_{173}$ and $conv2_{202}$ to the validation set (val. set), which consists of different species classes. Best viewed in electronic form.

Below each image patch is its deconvolution. In Fig. 6, we visualise the response of the selected filter units ($conv2_{151}$, $conv2_{139}$, $conv2_{173}$ and $conv2_{202}$) in layer 2. Although the top two image patches (Fig. 5(a)–(c)) show neurons activated on leaf blades, leaf tips and in certain regions of leaf blade respectively, based on the deconvolution on their validation set we notice a simple detection of gradient changes along the leaf structures at different orientations. Hence, these filters can be viewed as a set of gradient operators that extract dedicated edges or outlines of the leaf. On the other hand, for the channel $conv2_{202}$, the deconvolution on validation sets as well as the activation of the top two image patches (Fig. 5(d)) show similar effects, i.e. the neurons are highly activated at the surface of the leaf, covering the entire leaf area. The reason might be that the filters are focusing on the leaf color.

In Fig. 7, we visualise the response of the selected filter units ($conv3_4$, $conv3_{50}$, $conv3_{228}$ and $conv3_{265}$) in layer 3. In layer 3, we can observe more complex invariances than those of layer 2. For example, for the channel $conv3_4$, it can be seen that activation is located on divergent structures of the top two image patches (Fig. 5(e)). However, deconvolution on validation set shows that in all cases, the neurons are activated in the leaf base region. The reason might be that the filters are learning some kind of wave edge structure, which results in neuron activation being associated with cordate or cuneate-shaped leaf base features. For the channel $conv3_{50}$, the whole leaf boundary can be observed in the deconvolution of the validation set, showing the outline of an area of a leaf image. Hence, these filters can be regarded as a set of gradient operators that extract dedicated edges or leaf outlines. Next, for the channel $conv3_{228}$, arching shape outlines were activated in the top two image patches (Fig. 5(g)). This could be due to the filters capturing particular curving structures of the leaf as depicted in the deconvolution on validation set. For the channel $conv3_{265}$, neuron activation is located on the divergent structures (leaf veins) of the top two image patches (Fig. 5(h)), and the same response is observable in the deconvolution of the validation set.

In Fig. 8, we visualise the response of the selected filter units ($conv4_{373}$, $conv4_{170}$, $conv4_{148}$ and $conv4_{365}$) in layer 4. In layer 4, we observe mid-level semantic partial abstraction of leaf structures, where the features extracted have almost similar complexity levels to layer 3. For example: venation-like features are observed in the channel $conv4_{373}$ (Fig. 5(i)) based on the deconvolution result of the validation set; the neurons are not only activated on the divergent structures (secondary veins) but on the central veins (primary veins) as well. For the channel $conv4_{170}$, the selected filters are activated by the curvature of the lobed leaves, as shown in the deconvolution of the top two image patches (Fig. 5(j)). This can be interpreted as extraction of conjunctions of curvature features in certain orientations. On the other hand, for the chan-

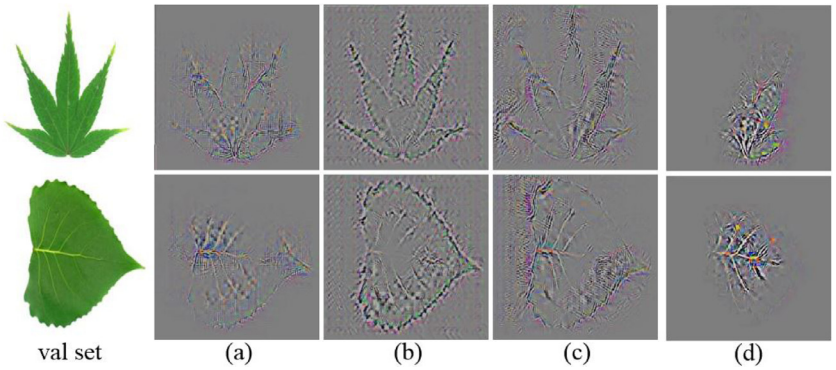

Fig. 7. Each column (a), (b), (c) and (d) depicts the deconvolution results of channels $conv3_4$, $conv3_{50}$, $conv3_{228}$ and $conv3_{265}$ to the val set. Best viewed in electronic form.

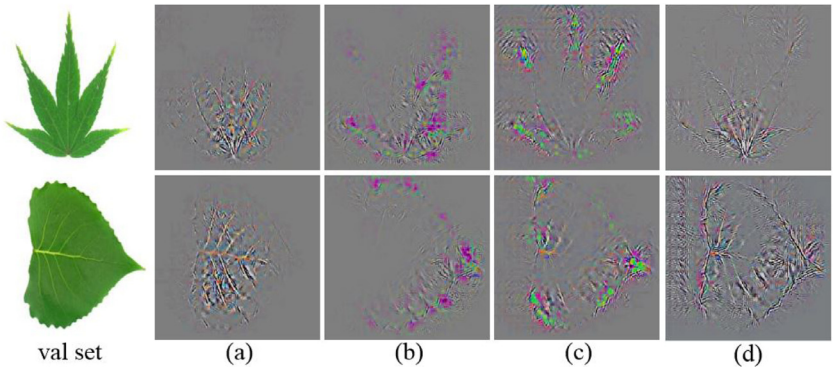

Fig. 8. Each column (a), (b), (c) and (d) depicts the deconvolution results of channels $conv4_{373}$, $conv4_{170}$, $conv4_{148}$ and $conv4_{365}$ to the val set. Best viewed in electronic form.

nel $conv4_{148}$, the deconvolution result of the validation set shows higher activation of neurons at sharp corners, especially the leaf tips. They are extracted based on filters that respond to leaf tips which taper into a long point, as depicted in the deconvolution of the top two image patches (Fig. 5(k)). For the channel $conv4_{365}$, we observe shape-like features appearing in some of the deconvolution results of the validation set, except in those leaves that have leaf or lobe tips tapering into a long point with toothed margins. The reason is because they are extracted based on filters that respond to corner conjunctions within a certain range of degree angles, as depicted in the deconvolution of the top two image patches (Fig. 5(l)).

From the filter visualization outcomes of layers 1–4, we observe a hierarchical transformation of features from low-level to mid-level abstraction. For example, from gradient changes to edges, then to the combination of edge-like divergent structures, and finally to mid-level abstraction of leaf-like entire venation structures. The higher level features build on the mid-level features while mid-level features build on the low-level features. Each is correlated, forming a robust feature representation for leaf images.

In layer 5, learned features show significant variation compared to previous layers, and are more class-specific. The learned filters do not show a similar response between species on the same leaf character. Here we show some examples of feature visualization in layer 5. Fig. 9 shows two groups of feature visualisations from channels $conv5_{32}$ (Fig. 9(a & b)) and $conv5_{168}$ (Fig. 9(c & d)) respectively. In each group, we examine the specificity of the features by comparing the activation regions of leaf images from different species classes (bounded by varying outline colors). In the leftmost figure, we show the top two image patches from the training set that caused the highest activations to the channel as well as the deconvolution results of the validation sets. Based on the deconvolutions of the validation set, we observe that neurons are mostly fired by specific kinds of leaf shape, leaf margin and venation. For example: pinnately veined leaves are activated, as shown in the validation set in Fig. 9(b). Next, neurons in $conv5_{168}$ are shown activated by leaf lobes with long, narrow blades, as shown in the validation set bounded by the blue outlines in Fig. 9(d). Although in each group both validation sets from different species classes have very similar leaf or lobe shapes, neurons are found to be activated only by higher-level features of the leaf structure representing particular characteristics of species. Therefore, unlike the general features discussed in previous layers, layer 5 features can be considered to be more specific for types of leaf structure which discriminate species classes.

5.3. Discussion

These findings deliver two important messages on leaf feature characterization. First, we observe a fruitful fact that features are learned in CNN transform from low-level to mid-level, and then finally to class-specific abstractions at the last convolutional layer. These findings fit with the hierarchical botanical definitions of leaf characters, which are described in great detail by Ellis and colleagues [44]. Orthodox taxonomic description of leaves proceeds hierarchically from the general (e.g. contour shape, the overall pattern of the major vein system,) to the particular (e.g. anterior and posterior lobes in lobed leaves, the two halves of the lobes, the secondary and tertiary vein systems, etc.). Each of these features may have several to many states at each hierarchical level. Secondly, the learned features are not merely constrained to shape, texture or color but also extend to specific kinds of leaf characters such as structural divisions, leaf tip, leaf base, margin types, etc. This shows that using DL approach, we are able to perceive the cognitive visual complexities for leaves, information which is often limited to the research community working on plant identification.

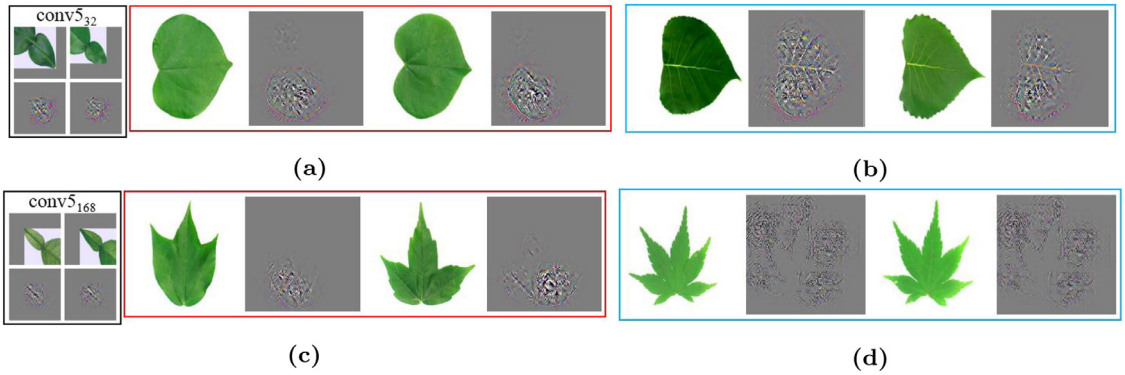

Fig. 9. Each row (a & b) and (c & d) depicts the deconvolution results of channels $conv5_{32}$ and $conv5_{168}$ respectively. Best viewed in electronic form.

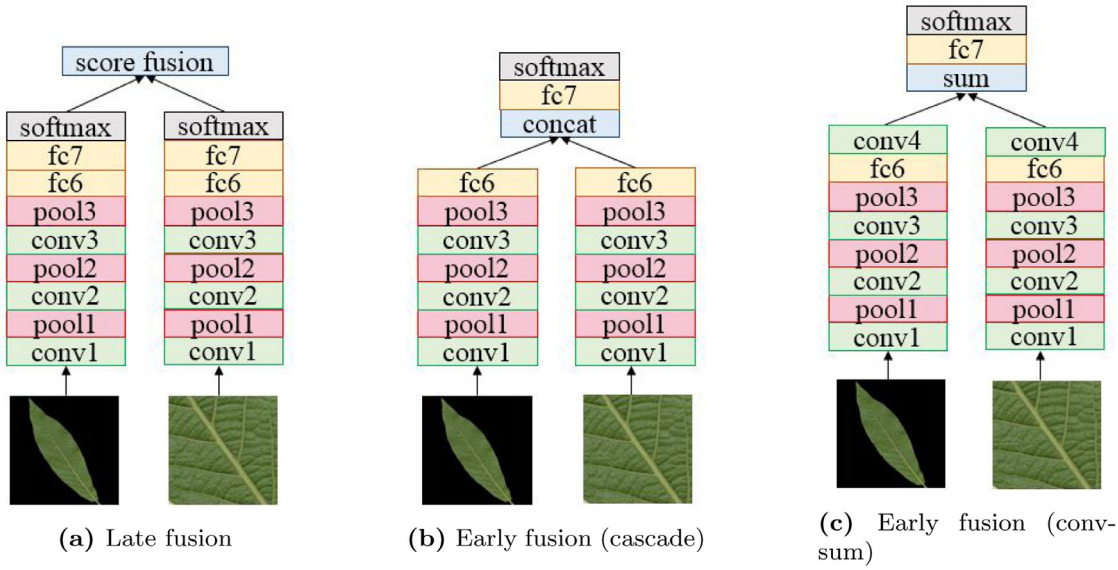

Fig. 10. Different types of fusion strategies.

6. Hybrid global-local leaf feature extraction

According to our preliminary experiments and visualisation outcomes in Sections 4 and 5, we gain an important intuition that CNN trained using whole leaves (D1) and leaf patches (D2) extract different levels of contextual information. As such, using the whole leaf image, we found the emergence of global features describing the holistic structure of leaf such as the shape, color, texture and margin, while using leaf patches we noticed that CNN tends to capture the local intrinsic patterns of venation. Importantly, these findings demonstrate that CNN trained with leaf patches is capable of recognising the relevant vein patterns and differentiating them among species without needing any manual segmentation or pre-processing [30] on veins.

Although venation is known to be the powerful alternative feature representation for leaf classification, others leaf features like shape and margin are usually used together with venation by plant taxonomists for classifying plant species. In this study, based on our discovery that CNN trained on different input data formats provides variants of contextual features of leaf, we design a new hybrid global-local feature extraction model for leaf data based on CNN approach. Instead of relying on either whole leaf data [31, 45–47] or solely venation [30,31] for species classification, we propose to combine information from two CNN networks, one global network trained upon the whole leaf data and another local network trained upon its corresponding leaf patches. We integrate them via different feature fusion strategies as illustrated in Fig. 10.

6.1. Approach

In this section, we consider different architectures for fusing both global and local information of leaf features: the *late* and *early fusion*. For late fusion, the fusion can be done at the corresponding softmax outputs after the pre-training of each CNN network, while for early fusion it can be carried out before class score computation, such as during the feature learning stage. We first introduce our new single network architecture and then discuss its extension to hybrid feature extraction based on different types of fusion strategy.

Single stream We design a new single stream CNN which comprises shorter depth layers for the purpose of testing out the stability of CNN as well as to evaluate the contribution of hybrid global-local features in species classification. Using shorthand notation, the full architecture is conv1(96,11,4) - pool1(96,3,2) - norm - conv2(256,5,1) - pool2(256,3,2) - norm - conv3(384,3,1) - pool3(384,3,2) - fc6(2048) - fc7(44), where pool$l$ or conv$l(c,k,s_t)$ indicates pooling or convolution in layer l with c number of channels computed by filters of size $k \times k$ with stride s_t. *norm* is the normalization layer defined in [32]. fc$l(v)$ is the lth fully connected layer with v nodes.

Late fusion To build a hybrid feature extraction model for leaf data, we devise our plant classification system accordingly, dividing a single architecture into two networks: a global and a local network, each trained on whole leaf images and their corresponding leaf patches respectively as shown in Fig. 10a. Here, we first

Table 5
Top-1 classification accuracy results of our proposed models. Note that, LF = late fusion, EF = early fusion, W = whole leaf, P = patches.

Model	Parameters(million)	Type	Number of Training data	
			W = 1,324, P = 3960	W = 2,288, P = 34,672
Finetuned AlexNet [32]	58	W	0.956	0.977
Finetuned AlexNet [32]	58	P	0.914	0.995
Single stream	30	W	0.915	–
Single stream	30	P	0.883	–
	60	LF (mav)	0.941	–
	60	LF (ave)	0.945	–
	60	EF (cascade)	0.955	–
	64	**EF (conv-sum)**	**0.963**	–

Table 6
Existing dataset examples.

Dataset	Quantity of images	Number of categories
MS COCO[51]	328k (2.5 million labeled instances)	91
Places2 [52]	8.3 million	365
Sport-1M[53]	1 million	487
Visual Genome QA [54]	1.7 million questions/answer pairs	–
ILSVRC 2010 [55]	1.4 million	1000
PlantClef2015 dataset [6]	**113,205**	**1000**

pre-train each network using its corresponding leaf data. During the validation phase, we combine both softmax outputs and compute the final class scores using fusion methods: average (ave) or max voting (mav).

Early fusion Early fusion models integrate both networks and jointly train them end-to-end with fused representation linked directly to the species classes via softmax layer. Note that, unlike the late fusion method, early fusion has fused representation learned conjointly with divided networks according to the species class labels. We consider two late fusion strategies: *cascade* (Fig. 10b) and *conv-sum* (Fig. 10c). In *cascade* fusion, f_{cas} stacks both fc6 layer's weight matrices across the feature channels, forming a cascaded matrix x_{cat} in concat layer: $\mathbf{x}_{cat} = f_{cas}(\mathbf{x_g}, \mathbf{x_o})$ where $\mathbf{x}_g, \mathbf{x}_o \in \mathbb{R}^{1\times1\times n}$ and resulting $\mathbf{x}_{cat} \in \mathbb{R}^{1\times1\times2n}$.

In *conv-sum* fusion, $\mathbf{x}_{cs} = fcs(\mathbf{x}^{gc}, \mathbf{x}^{oc})$, *fcs* first convolves each fc6 layer's weight matrix ($\mathbf{x}_g$, $\mathbf{x}_o$) with U numbers of filters $\mathbf{w}$ and biases $\mathbf{b}$: $\mathbf{x}^{jc} = \mathbf{x}_j * \mathbf{w}_j + \mathbf{b}_j$ where $j = \{g,o\}$ and each $\mathbf{w} \in \mathbb{R}^{1\times1\times n\times U}$ and $\mathbf{b} \in \mathbb{R}^U$, resulting in $\mathbf{x}^{gc}, \mathbf{x}^{oc} \in \mathbb{R}^{1\times1\times U}$. Both elements in $\mathbf{x}^{gc}$ and $\mathbf{x}^{oc}$ are then summed in the later stage. In our model, we set the number of filters U to 1000: $\mathbf{x}_{cs} = \sum_{r=1}^{U} x^{oc}_{1,1,r} + x^{gc}_{1,1,r}$ The difference between *cascade* and *conv-sum* fusion is that *conv-sum* fusion undergoes an additional convolution process to find out important features of each network before fusion. Next, during feature fusion, features summation is performed instead of concatenation to further amplify the correspondences of these features.

6.2. Experiments

In these experiments, we increase the difficulty of the classification problem by constraining the varieties of leaf data to be seen by the CNN during training. Hence, instead of considering all the existing training data, we left out some images for training. We adopt the MK dataset and compute a smaller training set of 57.9% of the whole leaf dataset (D1) to train on global network. From each leaf image, we randomly crop three leaf patches to train on the local network, accumulating a total of 3960 images, which is only 11.4% of leaf patch dataset (D2). In both networks, we maintain the size of testing set which is 528 images.

Instead of training both networks from random-initialised weight values, we transfer the weight matrices from the pre-trained model [32] and fine-tune it using our own leaf dataset. For the parameter setting in training, we employ fixed learning policy. We set the learning rate to 10^{-3}, and then decrease it by a factor of 10 when the validation set accuracy stops improving. The momentum is set to 0.9 and weight decay to 10^{-4}. In this experiment, we compute the top-1 classification accuracy as described in Section 4.3.

6.2.1. Results and discussion

Table 5 shows the comparison performance between single stream and the proposed hybrid feature extraction models. First of all, it is noticeable that classification performance is affected when we constrain the varieties of leaf data to be seen by CNN during training. This is clearly shown in the top-1 accuracy results of the finetuned AlexNet model. Classification performance of the network trained with all training sets (w = 2,288, P = 34,672) is obviously better compared to that trained on smaller subset of data (W = 1,324, P = 3,960). Next, although reducing CNN layer depth might affect feature discrimination power of a network, we found that combining both global and local leaf data is an alternative to boost the classification performance. Further analysis of EF and LF reveals that combining both features at the early stage is more beneficial as features are learned end-to-end, starting from before and after fusion. Moreover, we note that introducing a new feature subset learning stage before fusion at *conv-sum* can help to amplify the important features for each network, and with fusion through summation, we achieve the best accuracy of 0.963.

Based on all the facts that support the efficiency of leaf features learned using CNN for species identification, it now appears undeniable that CNN is a key tool to assist researchers to discover which the leaf features are most effective for plant species identification. Nevertheless, we come up against a common question that is often arises in the field of deep learning: how many convolutional layers are required in CNN to achieve the best optimization ability in modeling plant data? Is using only the AlexNet model sufficient? Based on numerous publications on object classification benchmarks, we observe a dramatic increase in depth for CNN in achieving the state-of-the-art result. For example: from 5 convolutional layers in AlexNet [32] to 16 in VGGNet [48], 21 in GoogleNet [49], and then to 164 in ResNet [50]. This conveys the important message that when the network goes deeper and deeper, its optimization capability can be further improved. However, deep CNN networks require very large amounts of training data. Table 6 shows examples of existing well-known datasets and their size as quantity of images. The biggest plant database that we have found is the PlantClef2015 dataset [6] which has only around 113,205 number of images. This is still far from matching the scale and variety of existing general major datasets for images [51,52,55], videos [53] or languages [54]. In addition, we can see that the PlantClef2015 dataset [6] has one of the largest number of object categories but the least number of images. For example, compared to the ILSVRC 2010 dataset [55], it has less than 10% of their total images but the same number of categories. Hence, to

efficiently train a deep architecture to recognize and learn features of plant images, much larger datasets are required, preferably with more than a million images and higher category variability to support future work by the research community in this area.

7. Conclusion

This paper investigated the use of deep learning to harvest discriminatory features from leaf images by learning, and apply them as classifiers for plant identification. Our experimental results demonstrate that learning the features using CNNs can provide better feature representations of leaf images as compared to using hand-crafted features. We also quantified the features that most efficiently represent the leaves for the purpose of species identification, using a DN approach. In the first experiment we show that venation structure is a very important feature for identification especially when shape feature alone is inadequate. This is verified by checking the global response of the filters in each convolution layer using the V1 strategy. We furthermore quantified the leaf image features by examining the local response of individual filters in each convolution layer. We observed a hierarchical transformation of features from low-level to high-level abstraction throughout the convolution layer, and these findings fit the hierarchical botanical definitions of leaf characters. Finally, we introduce new hybrid models, exploiting the correspondence of different contextual information of leaf features. We show experimentally that hybrid local-global features learned using DL can improve recognition performance compared to previous techniques.

Acknowledgement

This research is supported by the Fundamental Research Grant Scheme (FRGS) MoHE Grant FP070-2015A, from the Ministry of Education Malaysia; and we gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X GPU used for this research. S.H. Lee is supported by the Postgraduate Research (PPP) Grant PG007-2016A, from University of Malaya.

References

[1] R. Govaerts, How many species of seed plants are there? Taxon 50 (4) (2001) 1085–1090.
[2] H. Nagendra, D. Rocchini, High resolution satellite imagery for tropical biodiversity studies: the devil is in the detail, Biodivers. Conserv. 17 (14) (2008) 3431–3442.
[3] L. Qi, Q. Yang, G. Bao, Y. Xun, L. Zhang, A dynamic threshold segmentation algorithm for cucumber identification in greenhouse, in: International Congress on Image and Signal Processing, 2009, pp. 1–4.
[4] S. Zhang, Y. Lei, T. Dong, X.-P. Zhang, Label propagation based supervised locality projection analysis for plant leaf classification, Pattern Recognit. 46 (7) (2013) 1891–1897.
[5] S.G. Wu, F.S. Bao, E.Y. Xu, Y.-X. Wang, Y.-F. Chang, Q.-L. Xiang, A leaf recognition algorithm for plant classification using probabilistic neural network, in: IEEE International Symposium on Signal Processing and Information Technology, 2007, pp. 11–16.
[6] A. Joly, H. Goëau, H. Glotin, C. Spampinato, P. Bonnet, W.-P. Vellinga, R. Planqué, A. Rauber, S. Palazzo, B. Fisher, et al., LifeCLEF 2015: multimedia life species identification challenges, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction, Springer, 2015, pp. 462–483.
[7] N. Kumar, P.N. Belhumeur, A. Biswas, D.W. Jacobs, W.J. Kress, I.C. Lopez, J.V. Soares, Leafsnap: a computer vision system for automatic plant species identification, in: ECCV, Springer, 2012, pp. 502–516.
[8] D. Hall, C. McCool, F. Dayoub, N. Sunderhauf, B. Upcroft, Evaluation of features for leaf classification in challenging conditions, in: 2015 IEEE Winter Conference on Applications of Computer Vision, 2015, pp. 797–804.
[9] C. Kalyoncu, Ö. Toygar, Geometric leaf classification, Comput. Vision Image Understanding 133 (2015) 102–109.
[10] A. Kadir, L.E. Nugroho, A. Susanto, P.I. Santosa, Leaf classification using shape, color, and texture features, arXiv:1401.4447 (2013).
[11] J.C. Neto, G.E. Meyer, D.D. Jones, A.K. Samal, Plant species identification using elliptic fourier leaf shape analysis, Comput. Electron. Agric. 50 (2) (2006) 121–134.
[12] S. Mouine, I. Yahiaoui, A. Verroust-Blondet, Advanced shape context for plant species identification using leaf image retrieval, in: Proceedings of the 2nd ACM International Conference on Multimedia Retrieval, 2012, p. 49.
[13] X.-Y. Xiao, R. Hu, S.-W. Zhang, X.-F. Wang, Hog-based approach for leaf classification, in: Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence, Springer, 2010, pp. 149–155.
[14] Y. Naresh, H. Nagendraswamy, Classification of medicinal plants: an approach using modified LBP with symbolic representation, Neurocomputing 173 (2016) 1789–1797.
[15] Z. Tang, Y. Su, M.J. Er, F. Qi, L. Zhang, J. Zhou, A local binary pattern based texture descriptors for classification of tea leaves, Neurocomputing 168 (2015) 1011–1023.
[16] J.S. Cope, P. Remagnino, S. Barman, P. Wilkin, Plant texture classification using Gabor co-occurrences, in: International Symposium on Visual Computing, Springer, 2010, pp. 669–677.
[17] J. Charters, Z. Wang, Z. Chi, A.C. Tsoi, D.D. Feng, EAGLE: a novel descriptor for identifying plant species using leaf lamina vascular features, in: ICME-Workshop, 2014, pp. 1–6.
[18] M.G. Larese, R. Namías, R.M. Craviotto, M.R. Arango, C. Gallo, P.M. Granitto, Automatic classification of legumes using leaf vein image features, Pattern Recognit. 47 (1) (2014) 158–168.
[19] J. Chaki, R. Parekh, Plant leaf recognition using shape based features and neural network classifiers, Int. J. Adv. Comput. Sci. Appl. 2 (10) (2011).
[20] J.-X. Du, X.-F. Wang, G.-J. Zhang, Leaf shape based plant species recognition, Appl. Math. Comput. 185 (2) (2007) 883–893.
[21] A. Aakif, M.F. Khan, Automatic classification of plants based on their leaves, Biosyst. Eng. 139 (2015) 66–75.
[22] C. Zhao, S.S. Chan, W.-K. Cham, L. Chu, Plant identification using leaf shapes – a pattern counting approach, Pattern Recognit. 48 (10) (2015) 3203–3215.
[23] J.S. Cope, P. Remagnino, Classifying plant leaves from their margins using dynamic time warping, in: International Conference on Advanced Concepts for Intelligent Vision Systems, Springer, 2012, pp. 258–267.
[24] L.-H. Ma, Z.-Q. Zhao, J. Wang, ApLeafis: an android-based plant leaf identification system, in: International Conference on Intelligent Computing, 2013, pp. 106–111.
[25] A.R. Backes, O.M. Bruno, Plant leaf identification using multi-scale fractal dimension, in: International Conference on Image Analysis and Processing, Springer, 2009, pp. 143–150.
[26] M. Rashad, B. El-Desouky, M.S. Khawasik, Plants images classification based on textural features using combined classifier, Int. J. Comput. Sci. Inf. Technol. 3 (4) (2011) 93–100.
[27] A. Olsen, S. Han, B. Calvert, P. Ridd, O. Kenny, In situ leaf classification using histograms of oriented gradients, in: International Conference on Digital Image Computing, 2015, pp. 1–8.
[28] T. Beghin, J.S. Cope, P. Remagnino, S. Barman, Shape and texture based plant leaf classification, in: International Conference on Advanced Concepts for Intelligent Vision Systems, Springer, 2010, pp. 345–353.
[29] J. Chaki, R. Parekh, S. Bhattacharya, Plant leaf recognition using texture and shape features with neural classifiers, Pattern Recognit. Lett. 58 (2015) 61–68.
[30] G.L. Grinblat, L.C. Uzal, M.G. Larese, P.M. Granitto, Deep learning for plant identification using vein morphological patterns, Comput. Electron. Agric. 127 (2016) 418–424.
[31] S.H. Lee, C.S. Chan, P. Wilkin, P. Remagnino, Deep-plant: plant identification with convolutional neural networks, in: IEEE International Conference on Image Processing, 2015, pp. 452–456.
[32] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: NIPS, 2012, pp. 1097–1105.
[33] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, T. Darrell, Decaf: a deep convolutional activation feature for generic visual recognition, arXiv:1310.1531 (2013).
[34] C. Dong, C.C. Loy, K. He, X. Tang, Learning a deep convolutional network for image super-resolution, in: ECCV, Springer, 2014, pp. 184–199.
[35] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, Caffe: convolutional architecture for fast feature embedding, in: Proceedings of the 22nd ACM international conference on Multimedia, ACM, 2014, pp. 675–678.
[36] M.D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, in: ECCV, Springer, 2014, pp. 818–833.
[37] J. Yang, K. Yu, Y. Gong, T. Huang, Linear spatial pyramid matching using sparse coding for image classification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 1794–1801.
[38] A. Roth-Nebelsick, D. Uhl, V. Mosbrugger, H. Kerp, Evolution and function of leaf venation architecture: a review, Ann. Bot. 87 (5) (2001) 553–566.
[39] H. Candela, A. Martınez-Laborda, J. Luis Micol, Venation pattern formation in Arabidopsis thaliana vegetative leaves, Dev. Biol. 205 (1) (1999) 205–216.
[40] A. Runions, M. Fuhrer, B. Lane, P. Federl, A.-G. Rolland-Lagan, P. Prusinkiewicz, Modeling and visualization of leaf venation patterns, ACM Trans. Graph. 24 (3) (2005) 702–711.
[41] J. Clarke, S. Barman, P. Remagnino, K. Bailey, D. Kirkup, S. Mayo, P. Wilkin, Venation pattern analysis of leaf images, in: Advances in Visual Computing, Springer, 2006, pp. 427–436.
[42] J.S. Cope, P. Remagnino, S. Barman, P. Wilkin, The extraction of venation from leaf images by evolved vein classifiers and ant colony algorithms, in: Advanced Concepts for Intelligent Vision Systems, Springer, 2010, pp. 135–144.
[43] R.J. Mullen, D. Monekosso, S. Barman, P. Remagnino, P. Wilkin, Artificial ants to extract leaf outlines and primary venation patterns, in: Ant Colony Optimization and Swarm Intelligence, Springer, 2008, pp. 251–258.
[44] B. Ellis, D.C. Daly, L.J. Hickey, K.R. Johnson, J.D. Mitchell, P. Wilf, S.L. Wing, Manual of leaf architecture, 190, Cornell University Press Ithaca, 2009.

[45] A.K. Reyes, J.C. Caicedo, J.E. Camargo, Fine-tuning deep convolutional networks for plant recognition, in: Working Notes of CLEF 2015 Conference, 2015.
[46] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, D. Stefanovic, Deep neural networks based recognition of plant diseases by leaf image classification, Comput. Intell. Neurosci. (2016) 1–11.
[47] C. Ashley, D. Alexandre, D. Stewart, H. Gerard, L. Simon, M. John, Plant recognition: Bringing deep learning to iOS(2014).
[48] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv:1409.1556 (2014).
[49] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.
[50] K. Zhang, M. Sun, T.X. Han, X. Yuan, L. Guo, T. Liu, Residual networks of residual networks: multilevel residual networks, arXiv:1608.02908 (2016).
[51] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, Microsoft coco: common objects in context, in: ECCV, Springer, 2014, pp. 740–755.
[52] B. Zhou, A. Khosla, A. Lapedriza, A. Torralba, A. Oliva, Places:an image database for deep scene understanding, arXiv:1610.02055 (2016).
[53] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, L. Fei-Fei, Large-scale video classification with convolutional neural networks, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2014, pp. 1725–1732.
[54] R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L.-J. Li, D.A. Shamma, et al., Visual genome: connecting language and vision using crowdsourced dense image annotations, arXiv:1602.07332 (2016).
[55] A. Berg, J. Deng, L. Fei-Fei, Large scale visual recognition challenge 2010, 2010,

Sue Han Lee received Master degree in Electrical and Electronics Engineering from Shinshu University, Japan in 2014. She is currently pursuing the Ph.D. degree at the Faculty of Computer Science and Information Technology, University of Malaya, Malaysia. Her research interest is computer vision, with main focus on plant recognition.

Chee Seng Chan is a Senior Lecturer at the Faculty of Computer Science and Information Technology, University of Malaya, Malaysia. His research interests are computer vision and fuzzy set theory, particularly focus on image/video content analysis. Dr. Chan was the founding Chair for the IEEE Computational Intelligence Society (CIS) Malaysia chapter, the organising chair for the 3rd IAPR Asian Conference on Pattern Recognition (ACPR2015), and general chair for the IEEE Visual Communications and Image Processing (VCIP2013). He is a Senior Member of IEEE, a Chartered Engineer and a Member of IET.

Simon Joseph Mayo is a taxonomist based at the Royal Botanic Gardens Kew, UK, specializing in the taxonomy of the mainly tropical plant family Araceae. His areas of interest are morphometrics, species delimitation and botanical research and teaching in Brazil.

Paolo Remagnino is full professor in the Computer Science department at Kingston University, where he leads the multi-disciplinary Robot Vision Team. His research interests include image and video understanding, pattern recognition, machine, deep and manifold learning.