Deep Learning for Plant Species Classification Survey

Mary Sobha P.G.
MTech Scholar
Dept. of Computer Science & Engg.
Government Engineering College
Idukki
msobhapg@gmail.com

Mrs. Princy Ann Thomas
Associate Professor
Dept. of Computer Science & Engg.
Government Engineering College
Idukki
princyannsunil@gmail.com

***Abstract*—Protection of biodiversity is quite essential and for this purpose we should know more about the species. Identification of plant species by using conventional hand-crafted features is complex. It is difficult for non-experts to remember the specific botanical terms. The idea of automatic identification of plant species is approaching reality. Machine learning and deep learning play an important role in this matter. The deep learning Convolutional Neural Networks (CNN) can apply to extract the features from leaf images. CNN or machine learning classifiers can be used for the classification of plant species. The deep learning methods outperform all handcrafted methods. This survey is about the identification of plant species using deep learning methods.**

Keywords—plant species classification, convolutional neural networks, CNN, deep learning.

I. Introduction

Identification of plant species is usually done by Botanists and Plant Ecologists. This data is also needed by Architects, Farmers, Foresters, nature lovers, biologists, etc. There are lakhs of plant species in the world. The plant identification by conventional methods is difficult, time-consuming and frustrating for newcomers. Besides, due to high similarity between some plant species, it may be difficult to differentiate them easily. Many plants species face the problem of extinction. Endangered and non-endangered plant species need to be preserved and conserved in a proper way to reduce the risks of extinction. Hence, there is a need to develop an automated or computerized system to identify and classify the plants. Leaf is the common part of a plant available throughout the year for most of the plants. It's shape, textures, veins and colors are the main features that can be used to develop such an automated plant classification systems.

To increase human's ability to identify the plant species, recently in computer science research, lots of new techniques like image processing, pattern recognition, and machine learning are widely being used. In order to perform pattern recognition and classification, along with image processing machine learning (AI) techniques such as Artificial Neural network, Support Vector Machines, k-Nearest Neighbour and others are mainly used. Deep learning, a subfield of artificial intelligence (AI), is a popular and widely used technique that has been applied in various domains including biology, medical, computer vision and speech recognition. The deep learning methods have emerged as a promising alternative in plant recognition and it outperforms other hand-crafted methods in feature extraction.

This survey aims to study the deep learning networks and the research papers available on the topic of plant identification using leaf taxonomies & CNN.

II. Deep Learning

A deep learning neural network contains several nonlinear processing layers, working in parallel like biological nervous systems. An input layer, several hidden layers, and an output layer are the main parts of a deep neural network. All layers are interconnected by nodes, or neurons and each internal layer using the output of the previous layer as its input.

Deep learning is good for feature extraction and it is superior in providing deeper information of images than conventional machine learning techniques. Convolutional Neural Network (CNN), Deep Belief Network (DBN), Recurrent Neural Network (RNN), etc. are deep learning architecture.

Deep learning is a type of machine learning and do the feature extraction automatically and perform classification tasks directly from images, text, or sound. The term "deep" refers to the number of layers in the network—the more layers, the deeper the network.

The following technology enablers help to work with deep learning models easily.

- Availability of huge labeled datasets eg. ImageNet and PASCAL
- Increased computing power eg. GPUs
- Pretrained models built by experts eg. AlexNet

A deep learning architecture CNN is mainly used for feature extraction and classification of images.

III. Convolutional Neural Networks (CNN)

CNN is a type of feed-forward artificial neural network consisting of one or more convolutional layers which are then followed by one or more fully connected layers as in a standard Multi Layer perceptron (MLP).

A. Feature Extraction Layers

Three types of operations on data are mainly performed by this layer: Convolution, pooling, or rectified linear unit (ReLU).

1) Convolution: Input images are filtered using a set of convolutional filters and each filter activates some features from the images.
2) Pooling: The parameters of the learned network need to be simplified by using nonlinear downsampling.
3) ReLU: Used for faster and more effective training by mapping negative values to zero and maintaining positive values.

These three operations are repeated over tens or hundreds of layers, with each layer learning to detect different features.

B. Classification Layers

CNN architecture will do the classification using the feature maps learned in the feature extraction layer.

1) Fully Connected layer(FC): Outputs a vector of K dimensions where K is the number of classes that the network will be able to predict. Probabilities for each class of the images being classified is available in this vector.
2) Softmax function: The final layer of the CNN architecture uses a softmax function to provide the classification output.

If the number of input images are more, then the best result is obtained for deep learning architecture.

C. Transfer learning Of CNN models

Transfer learning is a deep learning method, here a model developed for one task is reused for another task by changing some layers of the network. For example, if Alexnet is used for identifying 40 new classes of objects instead of 1000 default classes, change the last fully connected layer for 40 classes.

IV. CNN Pre-trained Networks

CNN exists since 1980 but started to become popular when AlexNet was developed in 2012. AlexNet is a pre-trained CNN network prepared to identify 1000 classes of images. After AlexNet lots of such CNN networks are designed some of them are listed below. Table 1. Shows a comparison of some of the commonly used CNN models.

A. AlexNet

It is a CNN and Winner of ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012 designed by Alex Krizhevsky, and published with Ilya Sutskever, and Geoffrey E. Hinton, "Imagenet classification with deep convolutional neural networks"[12]. The architecture consists of eight layers five convolutional layers and three fully-connected layers. AlexNet uses Rectified Linear Units (ReLU) instead of the tanh function, which was standard at the time. AlexNet has 60 million parameters and to reduce overfitting Data Augmentation and Dropout methods were used.

B. GoogleNet

GoogleNet is designed by C. Szegedy et al. in their 2014 paper, “Going deeper with convolutions” [13]. It is a project from Google and the winner of ILSVRC 2014. GoogLeNet architecture is also known as the Inception Module. This module is based on several very small convolutions to drastically reduce the number of parameters. This architecture contains 22 layer deep CNN, the number of parameters reduced from 60 million to 4 million when compared with AlexNet. 1x1 convolutions are used in this architecture to reduce the dimensions inside the “inception module” and to add more non-linearity by having ReLU immediately after every 1x1 convolution.

C. VGGNet

Designed by Simonyan and Zisserman in their 2014 paper, “Very Deep Convolutional Networks for Large Scale Image Recognition” [14]. The network is 16 layers deep and can classify images into 1000 object categories, such as the keyboard, mouse, pencil, and many animals. Like AlexNet VGGNet has only 3x3 convolutions, but lots of filters. Trained on 4 GPUs for 2–3 weeks. VGGNet is used as a baseline feature extractor in many other applications and its weight configuration is publicly available. VGGNet consists of 138 million parameters, which is a little bit difficult to handle. VGGNet is using only 3×3 convolutional layers stacked on top of each other in increasing depth so it is simple. Reducing volume size is handled by max pooling. Two fully-connected layers, each with 4,096 nodes is then followed by a softmax classifier.

D. ResNet

Resnet is designed by He et.al. in their 2015 paper, "Deep Residual Learning for Image Recognition" [15]. The ResNet architecture has become a great work, demonstrating that extremely deep networks can be trained using standard SGD through the use of residual modules. ResNet architecture relies on micro-architecture modules also called "network-in-network architectures". The term micro-architecture refers to the set of "building blocks" used to construct the network. A collection of micro-architecture building blocks along with standard CONV, POOL, etc. layers leads to the macro-architecture. Resnet has 152 layers still having lower complexity than VGGNet.

E. MobileNet

MobileNet is an architecture that is more suitable for mobile and embedded based vision applications where there is a lack of computing power. This architecture was proposed by Google. MobileNets are based on a streamlined architecture that uses depth-wise separable convolutions to build light weight deep neural networks [16]. The normal convolution is replaced by depthwise convolution followed by pointwise convolution which is called depthwise separable convolution. There are no pooling layers in between these depthwise separable blocks. The convolution layers are followed by batch normalization and the activation function used by MobileNet is ReLU6.

TABLE 1. Comparison of some CNN models

Year	Name of CNN	No of Parameters	Details of Layers
2012	AlexNet	60 million	8 layer
2014	GoogLeNet	4 million	22 Layers
2014	VGGNet	138 million	VGG16 - 16 layers VGG19 - 19 layers
2015	ResNet	ResNet50 - 25.6M ResNet101 - 44.5M	152 layers
2017	MobileNet	MobileNet V1 - 4.24M MobileNet V2 - 3.47M	28 layers

V. Related Work

A. Plant Identification using Conventional Methods

Tristen Munisami et.al. designed a plant species identification system using conventional methods for feature selection and machine learning techniques for classification [1]. A series of image processing techniques are used for preprocessing the image such as Rotation, gray scaling, thresholding, opening operations, inverse threshold, edge extraction, and edge filtering. The following features are extracted from the image such as Convex hull information, morphological information, Distance maps, centroid radial map, color histogram. K-nearest neighbor classifier is used for classification. Their own leaves dataset named Folio was used with 32 species of plants and 20 images for each species. 87.2% accuracy obtained for the Folio dataset and 91.1% accuracy obtained for the Flavia dataset.

Olfa Mzoughi et.al. designed a "Multiple leaflets-based identification approach dedicated to compound leaves" [2]. Here only machine learning methods used for classification. This paper is included in the review just to mention the compound leaves of plants. According to the shapes, leaves of the plant can be classified as simple or compound. In compound-leaves can see a collection of simple leaf-like structures called leaflets. In compound leaf identification considers this arrangement of leaflets and their special shapes. The identification process has 3 steps. (1) The first step is the leaflets extraction. (2) The second step is the local description of leaflets. In this step, test four classical texture descriptors (Hough, Fourier, Leoh, Eoh). (3) The third step, by using three fusion algorithms IRP, LO, and distInc. do the fusion of ranking lists obtained for each leaflet queries. Experiments were performed on compound leaves of the Pl@ntLeaves Scan pictures dataset for the different configurations of descriptors and fusion algorithms. They have shown an improvement in the classification rates with different values of the KNN classifier.

B. Plant Identification using CNN

Guillermo L. Grinblat et.al. developed a plant identification system using deep learning for feature extraction [3]. More attention has been paid to vein morphological patterns as a leaf fingerprint. A clear correlation has been established between vein characteristics and some properties of the leaf such as damage and drought tolerance. This suggests that vein morphology carries information suitable for plant classification when shape, color or texture differences are unobservable, as in the case of trying to separate different cultivars from the same species. The CNN avoids the use of handcrafted feature extractors from the pipeline. This deep learning approach remarkably improves the accuracy of the referred pipeline. Better accuracy obtained in the deep learning model compared with the traditional machine learning model.

Sue Han Leea et.al. developed a system to study how deep learning extracts and learns leaf features for plant classification [4]. 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 species identification, using a Deconvolutional Network (DN) approach. The experiment shows that venation structure is a very important feature for identification especially when shape feature alone is inadequate. A new hybrid global-local feature extraction model is designed for leaf data using CNN. Previous models used either whole leaf data or solely venation for species classification. Here a global CNN network trained upon the whole leaf data and another local CNN network trained upon its corresponding leaf patches are combined to form a new network. Integrate these 2 CNNs using different feature fusion strategies. Hybrid local-global features learned using DN can improve recognition performance compared to previous techniques.

Aravind Krishnaswamy Rangarajan et.al. designed a model to identify 6 tomato plant diseases from 7 classes using CNN models [5]. The transfer learning approach is used in this study with AlexNet and the VGG16 net. The accuracy of classification obtained using AlexNet and VGG16 net was 97.49% and 97.23% respectively for 13,262 images. VGG16 net outperforms in almost all cases compared to the AlexNet and both models got better accuracy when 373 (maximum number of images in one of the classes) images batch size were used.

XIHAI ZHANG et.al. designed 2 improved deep convolutional neural networks models, GoogLeNet and Cifar10 for identifying 9 types of maize leaves [6]. The improvements are done by adjusting the model parameters, changing the pooling combinations, adding the dropout operation and rectified linear unit (Relu) function, and reducing the number of classifiers. The modified models of GoogLeNet and Cifar10 achieve high accuracy (98.9%, 98.8% than their original model.

Ali Beikmohammadi et.al. designed a CNN network from MobileNet using transfer learning for Flavia and Leafsnap datasets [7]. MobileNet is trained for the new task as a feature extractor machine and logistic regression classifier trained on the target dataset. Transfer learning from pre-trained MobileNet's large dataset to a limited botanical dataset in plant recognition is well done. In the new MobileNet dataset Flavia with 32 classes and Leafsnap with 184 classes, achieving an accuracy of 99.6% and 90.54%, respectively. The performance of MobileNet after the large change in the number of classes has got better results than methods based on the hand-crafted features and other methods based on deep learning in terms of memory and precision.

Jing Hu et.al. designed a multi-scale fusion convolutional neural network (MSF-CNN) for plant leaf recognition [8]. The deep feature learning architecture with three basic units, CBR, MP (max pooling) and AP (average pooling) units. In MSF-CNN, an input image is down-sampled into multiple low-resolution images and these input images are step-by-step fed into the MSF-CNN architecture to learn discriminative features at different depths. Feature fusion is done by concatenation operation and got a better performance. The MSF-CNN method can obtain a better generalization ability.

Jing Wei Tan et.al. designed a new CNN model named D-Leaf Plant Species Classification using Leaf Vein Morphometric [9]. All images were converted to RGB to grayscale and then Sobel was used to segmenting out the region of interest (ROI) from the images. Segmented images are skeletonized to get a clean vein architecture. For feature extraction three CNN models - pre-trained AlexNet, fine-tuned AlexNet and a new model D-Leaf were used. All these three models with ANN classifiers got the best accuracy than other classifiers. A comparison between AlexNet and D-Leaf models showed that AlexNet is more complicated than D-Leaf and a system with more convolutional layers needs a longer execution time.

Voncarlos M. Araujo et.al. designed a method for fine-grained plant classification given the leaf image in which complementary global and patch-based features are combined at each hierarchical level (genus and species) by pre-trained and fine-tuned CNNs [10]. This model adopted hierarchical classification strategies for plant identification. The data augmentation techniques can be used to face the problem of plant classes with very few samples for training in the available imbalanced dataset. The hierarchical approach (genus-to-species) achieved the final average classification score of 86.44%.

PENG JIANG et.al. designed a new apple leaf disease detection model that uses deep learning with CNNs [11]. A new CNN model named INAR-SSD was designed by using a Googlenet Inception module and Rainbow concatenation SSD (R-SSD). The apple leaf disease dataset (ALDD) is composed of laboratory images and complex images under real field conditions are first constructed via data augmentation and image annotation technologies. A real-time detection model that is based on the single-shot multibox detector (SSD) for apple leaf diseases is proposed. VGGNet is modified to obtain the new basic pre-network, namely, VGG-INCEP (VGGNet with the Inception module) by introducing the GoogLeNet Inception module to improve the extraction performance for multiscale disease spots. For rainbow concatenation, pooling and deconvolution are utilized simultaneously to integrate context and fuse features of the feature pyramid at the back of the SSD. Because of R-SSD small diseased object detection performance is realized. INAR-SSD model realizes a detection performance of 78.80% mAP on ALDD, with a high-detection speed of 23.13 FPS.

In real-time images predicting the location of the object along with the class is called object detection. R-CNN, SSD, D-SSD, R-SSD, etc are some CNN based detection models. TABLE 2 shows the summary of the papers reviewed in this survey with advantages of each work and the method used for feature extraction and classification.

TABLE 2. Summary of related works

Ref.No	Year	Feature Extraction & Classification	Advantages	Datasets & Accuracy
[1]	2015	Conventional Methods- Convex hull, morphological, Distance maps, colour histogram etc. Classifier: KNN	Conventional methods used for feature extraction.	Folio (Own dataset) Folio - 87.2% Flavia - 91.1%
[2]	2014	Conventional methods classifier : KNN classifier	Compound Leaf classification	Pl@ntLeaves scan database Improvement in accuracy
[3]	2016	Preprocessed image + Deep learning- CNN model	Feature extraction by CNN and classification only by using vein morphological patterns	866 leaf images provided by INTA. Better accuracy in CNN model than PDA model
[4]	2017	CNN model for feature extraction and DN for visualisation of features. Hybrid global-local model Classifiers: MLP and SVM	CNNs can provide better feature representations of leaf images as compared to using hand-crafted features	Malayakew dataset MLP - 97.7% SVM - 98.1%
[5]	2018	CNN - transfer learned models of AlexNet and VGG16 net. Modified the number of images, mini batch sizes weight and bias learning rate.	VGG16net with minibatch size is increased accuracy decreased. Alexnet wt & bias learning rate above 30 accuracy increased	PlantVillage dataset AlexNet 97.49% GG16 net 97.23%
[6]	2018	Modified GoogLeNet and Cifar10 - CNN models.	Better Accuracy than Original GoogLeNet and Cifar10 models	500 Images from Plant Village dataset and Google webpages GoogLeNet - 98.9%, Cifar10- 98.8%
[7]	2018	CNN - transfer learned model of MobileNet Classifier: Regression	Transfer learned MobileNet for fewer classes got better performance than new models	Flavia, Leafsnap Flavia 99.6% Leafsnap 90.54%
[8]	2018	Multi-scale fusion convolutional neural network (MSF-CNN)	MSF-CNN architecture to learn discriminative features at different depths.	MalayaKew MSF-CNN 99.05
[9]	2018	Deep learning- CNN-3 models - pre-trained AlexNet, fine-tuned AlexNet and D-Leaf. Classifiers : SVM, ANN, k-NN, NB and CNN	D-Leaf model needs less execution time compared with AlexNet.	MalayaKew, Flavia and Swedish Leaf Dataset. D-Leaf - 94.88% AlexNet -93.26% fine-tuned AlexNet-95.54%
[10]	2018	2 CNNs trained on global and patch-based features for genus classification	hierarchical classification strategy is better than flat classification strategy. (family, genus, and species).	ImageCLEF 2015 86.44%
[11]	2019	SSD with GoogLeNet Inception module & Rainbow concatenation to improve the extraction performance	A real-time detection model based on the single-shot multibox detector (SSD).	Detection performance of 78.80% mAP (mean Average Precision)

VI. Conclusion

This Survey is focused on automated plant recognition systems based on deep learning convolutional networks. When started by using deep learning networks, a very big change occurred in the identification process and obtained very good accuracy compared with the machine learning techniques used before. Deep learning networks are good if we have a large dataset, otherwise machine learning technique seems to be good. In deep learning, feature extraction of images is done automatically and it provides deeper information of images. A deep learning architecture Convolutional Neural Network(CNN) is good for feature extraction and classification of images. Many pre-trained CNN models like AlexNet, GoogLeNet, MobileNet, etc are available and are prepared to identify some classes of objects. These pre-trained models can be used for our requirements by utilising transfer learning or fine-tuning the network according to our needs. Our own CNN models can also be designed for identification purposes.

In this survey, it is found that some of the papers used transfer learned or modified pre-trained networks. Whereas, in few papers, specially designed CNNs, hybrid models of CNNs, and fusion models of CNNs are used. Some CNN models increased the recognition accuracy and some of them reduced execution time.

In most of the papers, plant identification using simple leaves are only considered and compound leaves are not considered. There are 2 types of compound leaves, i.e., palmate and pinnate. From this pinnate compound leaves structure is different from simple leaves.

There are lakhs of plant species available in the world and here most of the researchers are dealing with very few species. Moreover, the learning time is more for deep learning. Real-time detection of plant species by using images in the real background should be considered in future studies. Object detection by using SSD, RSSD, etc. can also play a big role in this real-time image detection.

VII. References

[1] Trishen Munisami, Mahess Ramsurn, Somveer Kishnah, Sameerchand Pudaruth, “Plant Leaf Recognition using Shape features and colour Histogram with K-nearest Neighbour classifiers”, Second International symposium on Computer Vision and Internet (VisionNet’15), Elsevier Procedia Computer Science 58(2015) 740-747.-2015

[2] Olfa Mzoughi, Itheri Yahiaoui, Nozha Boujemaa, Ezzeddine Zagrouba, ”Multiple leaflets-based identification approach for compound leaf species”, 2014

[3] Guillermo L. Grinblat, Lucas C. Uzal ⇑ , Mónica G. Larese, Pablo M. Granitto, “Deep learning for plant identification using vein morphological patterns”, Elsevier-2016

[4] Sue Han Leea , Chee Seng Chana,∗ , Simon Joseph Mayo b , Paolo Remagninoc, “How deep learning extracts and learns leaf features for plant classification”, Elsevier, 2017

[5] Aravind Krishnaswamy Rangarajan, Raja Purushothaman*, Aniirudh Ramesh, ”Tomato crop disease classification using pre-trained deep learning algorithm”, Elsevier - 2018.

[6] XIHAI ZHANG , (Member, IEEE), YUE QIAO , FANFENG MENG, CHENGGUO FAN , AND MINGMING ZHANG, “Identification of Maize Leaf Diseases Using Improved Deep Convolutional Neural Networks” IEEEAccess

[7] Ali Beikmohammadi, Karim Faez, “Leaf Classification for Plant Recognition with Deep Transfer Learning” ,4th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS),IEE,2018.

[8] Jing Hu, Zhibo Chen∗ , Meng Yang, Rongguo Zhang and Yaji Cui, “A Multi-Scale Fusion Convolutional Neural Network for Plant Leaf Recognition” IEEE, 2018

[9] Jing Wei Tan, Siow-Wee Chang, Sameem Abdul-Kareem, Hwa Jen Yap, Kien-Thai Yong, “Deep Learning for Plant Species Classification using Leaf Vein Morphometric”, IEEE, 2018

[10] Voncarlos M. Araujo ´ ∗, Alceu S. Britto Jr∗§, Andre L. Brun ´ ∗, Luiz E. S. Oliveira†, Alessandro L. Koerich, “Fine-Grained Hierarchical Classification of Plant Leaf Images using Fusion of Deep Models”, IEEE conf, 2018

[11] PENG JIANG , YUEHAN CHEN , BIN LIU, DONGJIAN HE, AND CHUNQUAN LIANG, “Real-Time Detection of Apple Leaf Diseases Using Deep Learning Approach Based on Improved Convolutional Neural Networks”, IEEEAccess, 2019

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in neural information processing systems, 2012, pp. 1097-1105.

[13] C. Szegedy et al., “Going deeper with convolutions,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1-9.

[14] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.

[15] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.

[16] A. G. Howard et al., “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” arXiv preprint arXiv:1704.04861, 2017.

[17] Jana Wäldchen1, Patrick Mäder2, “Plant Species Identification Using Computer Vision Techniques: A Systematic Literature Review”, 2018

[18] W. Liu et al., ‘‘SSD: Single shot multibox detector,’’ in Proc. Eur. Conf. Comput. Vis., 2016, pp. 21–37.