



Article Plant Leaves Recognition Based on a Hierarchical One-Class Learning Scheme with Convolutional Auto-Encoder and Siamese Neural Network

Lamis Hamrouni^{1,*}, Mohammed Lamine Kherfi², Oussama Aiadi³ and Abdellah Benbelghit^{3,*}

- ¹ Laboratoire de Génie Electrique (LAGE), University of Kasdi Merbah Ouargla, PB 511, Ouargla 30000, Algeria
- ² LAMIA Laboratory, University of Québec à Trois-Riviéres, Trois-Riviéres, QC G8Z 4M3, Canada; Mohammedlamine.Kherfi@uqtr.ca
- ³ LINATI Laboratory, University of Kasdi Merbah Ouargla, PB 511, Ouargla 30000, Algeria; aiadi.oussama@univ-ouargla.dz
- * Correspondence: lamis0215@gmail.com or hamrouni.lamis@univ-ouargla.dz (L.H.); accessbenbelghit.hassi@gmail.com or benbelgith.abdallah@univ-ouargla.dz (A.B.)

Abstract: In this paper, we propose a novel method for plant leaves recognition by incorporating an unsupervised convolutional auto-encoder (CAE) and Siamese neural network in a unified framework by considering Siamese as an alternative to the conventional loss of CAE. Rather than the conventional exploitation of CAE and Siamese, in our case we have proposed to extend CAE for a novel supervised scenario by considering it as one-class learning classifier. For each class, CAE is trained to reconstruct its positive and negative examples and Siamese is trained to distinguish the similarity and the dissimilarity of the obtained examples. On the contrary and asymmetric to the related hierarchical classification scheme that doesn't require such a pre-knowledge and can be employed by non-experts automatically. We cluster the dataset to assemble similar classes together. A test image is first assigned to the nearest cluster, then matched to one class from the classes that fall under the determined cluster using our novel one-class learning classifier. The proposed method has been evaluated on the ImageCLEF2012 dataset. Experimental results have proved the superiority of our method compared to several state-of-the art methods.

Keywords: plant leaves classification; hierarchical classification; Siamese neural network; convolutional auto-encoder; one class learning

1. Introduction

Plants have a significant impact on human life and development; without them, there will be no existence of the earth's ecology [1]. Plants play a decisive role in providing oxygen, clean air, food, etc. Additionally, they contribute to several tasks of scientists from different domains such as agriculture, medicine, and environmental fields.

Traditionally, botanists classify plants manually by using molecular biology and cellular features of leaves. Nevertheless, with the huge number of plants that exist on the earth classification through experts and botanists is subjective and requires much effort from experts. Besides, this process is too expensive in terms of time and effort.

On the contrary, with the development of computer software and hardware, mobile devices, and image processing. Automatically performing such a task, using machine learning techniques, is rapid, inexpensive, and accurate as well. Automatic plant identification has become a hot research topic in recent years.

Plants can be classified using their organs such as leaves, stems, fruits [2,3], or flowers. Nevertheless, the leaf is the most adopted part for recognition purposes since it carries out the plant's inherent properties and it is available all the seasons, contrary to the other parts. In addition, the leaf flatness makes it easy to represent it by machine.



Citation: Hamrouni, L.; Kherfi, M.L.; Aiadi, O.; Benbelghit, A. Plant Leaves Recognition Based on a Hierarchical One-Class Learning Scheme with Convolutional Auto-Encoder and Siamese Neural Network. *Symmetry* 2021, *13*, 1705. https://doi.org/ 10.3390/sym13091705

Academic Editors: Basil Papadopoulos and Jan Awrejcewicz

Received: 29 July 2021 Accepted: 10 September 2021 Published: 15 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). In the literature on this subject, many researchers have attempted to put forward systems that are capable of automatically identifying plant species based on leaf images. Some approaches have resorted to handcrafted features, such as shape, for describing leaves [4,5] and texture for describing the veins [6,7] or by the combinations of both [8]. Some others have tackled the problem by introducing special discriminative plant leaf features [9] (or domain knowledge) that are based on botanical characteristics. In general, handcrafted features are defined manually and extracted through instructed algorithms. However, as a matter of fact, this process is complex and requires changes and recalculation for each problem or data set.

Recently, with the impressive performance of deep learning, neural networks can learn essential characteristics directly and automatically from raw images. Deep learning approaches have been successfully applied in solving different issues including facial expression recognition [10], medicine classification [11], and plant identification [12].

Despite the considerable efforts that have been undertaken by researchers, automatic plant identification from leaf images is still an open issue, and there is room for improvement.

In this paper, we put forward a novel hierarchical method for automatic plant leaves recognition based on a novel classifier that consists of incorporating Siamese as an alternative to traditional loss within a convolutional auto-encoder (CAE). We consider a one-class learning strategy, in which a CAE is trained for each class. For a test image from class #N, the loss yielded by the class N's auto-encoder is supposed to be much smaller compared to the losses produced for the other classes. However, this raises another issue since the CAE trained on complicated leaf images is capable of perfectly reconstructing those relatively easier images from different classes. To handle such an issue, we propose training a Siamese on top of each CAE (i.e., the CAE of each class). This one-shot learning strategy (i.e., SCNN) is considered as intelligent loss, which is an alternative to the conventional CAE loss. Siamese is integrated to learn symmetric/asymmetric between images belonging to the same class and those from different classes, respectively.

Plants are organized in a hierarchical order (i.e., family, genus, and species). According to the literature, hierarchical plant classification applied by the relevant methods [13,14] consists in assigning test images (first to the coarse classes and then to the fine classes) by progressing through the plant's hierarchy (i.e., genus, and species). Nevertheless, this process requires pre-knowledge of the dataset being classified, which is actually difficult to do, especially for non-experts. In this paper, we propose a hierarchical classification scheme that doesn't require pre-knowledge, and which can be used by non-experts. Our scheme consists in clustering the entire dataset to gather symmetric classes together. A test image is first assigned to the most suitable cluster using a clustering algorithm and then matched to one class from the classes that fall under the detected cluster by using our novel one class learning classifier (i.e., CAE based on Siamese as an alternative loss).

The evaluation of the proposed method was carried out on a well-known dataset, namely ImageCLEF2012. Experimental results have demonstrated the efficiency of our method and a noteworthy performance has been reached compared to certain other methods. The remainder of this paper is organized as follows. A brief review of related work is presented in Section 2. In Section 3, we describe the details of the proposed method. The experimental results are presented in Section 4. Finally, we draw some conclusions for future work.

2. Related Work

In recent years, a lot of effort has been made to achieve more reliability in automatic leaf recognition. From the literature, two main approaches have been considered: plant classification based on hand-designed features using a classifier and deep learning strategies. For the first one, features are chosen manually and extracted through instructed algorithms, then only a subset of the most discriminant features are considered the obtained handcrafted features are used to train classifiers (SVM, NN, NB, etc. [15]). For the

second one (i.e., deep learning), it can learn discriminative characteristics from the raw images and recognize them automatically.

As instance for the hand crafted features, Hu et al. [16] proposed a contour-based shape descriptor named the multi-scale distance matrix (MDM) for fast plant leaf recognition. They used the matrix for pairwise distances between points sampled on the boundary of a leaf to capture the geometric structure of the shape, and 1-NN in the classification stage. In [17], Wang et al. proposed a method that uses a multi-scale arch height (MARCH), where the hierarchical arch height features at the K-scale are extracted from each contour point to capture concave and convex characteristics. This method provides a coarse-to-fine shape description of the leaf. The recognition rate was calculated using the 1-Nearest-Neighbor classifier, and a prototype system for online plant leaf identification was developed to be used on a mobile platform. The authors in [18] proposed to represent the leaf contour using two matrices. The first one is the sign matrix to extract the convex and concave features, and the second one is the triangle center distance to extract the spatial properties of the contour; the 1-NN is used for recognition.

Although leaf shape may be adequate to distinguish between some species, the shapes of others may be highly symmetrical, making differentiation difficult. Such a problem could be solved by taking additional leaf features such as texture and veins. According to [15], GLCM and Gabor wavelet are the most commonly used texture features. Typically, as in [9], wavelets have been used to decompose images, fractals to extract features, and artificial neural network to classify leaf images. Ghasab et al. [19] used texture features derived from GLCM, namely contrast, correlation, energy, homogeneity, and entropy, and combined them with shape, color, and vein features. In [20], Kadir et al. built foliage plant identification systems. Zernike moments were combined with other features (namely geometric features, color moments, and gray-level co-occurrence matrix (GLCM)). The results show that Zernike Moments have a prospect as features in leaf identification systems once they are combined with other features. In [21], a modified local binary pattern was proposed to extract texture features, and a simple nearest neighbor classifier was performed for classification, to decrease the intra-class variation the clustering was exploited in order to group symmetric leaf samples; the results prove that considering texture features alone is not sufficient. In [22], the authors propose to classify plant species using 19 leaf venation features using a support vector machine (SVM) with an RBF kernel. In [23], the authors propose to identify plant leaf based on visual features using different artificial intelligence techniques such as artificial neural networks, the naive Bayes algorithm, the random forest algorithm, the K-nearest neighbor (KNN), and the support vector machine (SVM). The best results were carried by SVM. In [24], the authors propose morphological features and the support vector machine (SVM) with an adaptive boosting technique to classify plants.

Despite the effectiveness of the handcrafted features in the plant classification system, such features are limited to specific conditions, if the characteristics of the images change (e.g., over space or time), then the performance of these algorithms significantly decreases. In the last few years, to overcome the drawbacks of existing approaches, deep learning methods have proved to demonstrate significant success in several plant identification systems.

For instance, in [25], the authors proposed a CNN model for plant leaf classification (Leaf-Net). The model was carried on three public datasets. The results prove that CNN outperformed the hand-crafted method. In [26], the authors recognized leaf image at different scales. Images are first down-sampled into multiples low resolution images. Then, in order to learn different characteristics in various layers, the MSF-CNN is proposed. The final feature is obtained by fusing all the last layer information. The classification is performed using either a support vector machine (SVM) or multi-layer perceptron (MLP) classifier. In [27], Lee et al. designed a hybrid feature extraction models for plant identification based on de-convolution neural network. They attempted to analyze how CNN learn directly features from the raw representations of an input image. Their main conclusion was that veins are the best representative features compared to those of outline shape. Ghazi et al. [28] analyzed the influence of different parameters, such as batch size

and number of iterations, on the performance of the different deep learning architectures, including Google-Net, Alex-Net, and VGG-Net. They revealed that the number of iterations was the most significant factor that affects fine-tuning performance, whereas data augmentation comes in the second place. In [29], the authors proposed to identify leaf species by fine-tuning the Alex-Net.

In [13], the authors present a fine-grained plant leaf classification method based on the fusion of deep models. The basic idea consists of the adoption of hierarchical classification strategies by using two levels of CNN. In the first level, global features are extracted, while in the second one, local features are considered. The fusion of the hierarchical levels is conducted using a coarse-to-fine strategy (i.e., the predicted coarse categories (i.e., genus) are used to define which subordinate category will be evaluated during the fine prediction (i.e., species)). Similarly, in [14], the authors also proposed two representations as in the previous work, albeit by considering Siamese at each level to overcome unbalanced and scalable problems. In [30] they performed a comparison between Siamese and CNN for plant species identification with small datasets. Their conclusion is that the Siamese performed better than CNN in terms of lower computational cost and can generalize better than CNN.

Despite the improvement brought by deep learning in plant recognition, most methods exploit it as a feature extractor whilst it has several properties that can be introduced to improve and give reliable results. Furthermore, according to the literature, most methods treat plant identification as a flat classification problem, whereas plant hierarchical organization may serve to accelerate and facilitate the identification process as well as reduce the problem of inter-species.

3. Proposed Method

In this section, we give details on the proposed method for plant leaves recognition. In this work, we propose a hierarchical plant classification system based on one-class learning scheme with convolutional auto-encoder and Siamese neural network. The hierarchy of our system consists mainly of two steps (clustering and classification). Figure 1 presents a Hierarchical classification scheme followed by the proposed method.



Figure 1. Hierarchical classification scheme followed by the proposed method (stage1: clustering, stage2: classification using our novel method (classifier), TRC: the trained novel classifier, C: the classes).

Through this section, we first show the general pipeline of our novel classification scheme that is based on one-class learning strategy based on a convolutional auto-encoder using a Siamese neural network as an intelligent loss. Then, we present details on each step, and at the end of this section, we will provide the general strategy of our hierarchical scheme.

3.1. General Pipeline of the Novel Classifier

In the proposed plant classification system, each image I_i is labeled with a label from the set = $\{i/i = 1, ..., n\}$, where *n* stands for the number of classes. For each class c_i we design a convolutional auto-encoder, denoted as CAE_i . Images within class c_i are firstly fed to the encoder to generate the latent representation termed as R ($I_j \rightarrow R$), then the decoder reconstruct the code *R* to produce the reconstructed image I'_j ($R \rightarrow I'_j$). On the top of each CAE_i , a Siamese is integrated as an indicator for the class to which an image sample belongs. After training CAE_i for each class, *Siamese_i* is trained using positive examples (represent images I_i and its reconstructions I'_i generated by the trained CAE_i) and negative examples (represent images I_j and its reconstructions I'_j generated by the trained $CAE_i/(j \neq i)$). The Siamese is an efficient alternative of the conventional loss of CAE_i . This one-class strategy is repeated for all of the remaining classes.

To sum up, hereafter, we summarize the steps of our method

• In the first step, the unsupervised CAE of each class c_i (from the set of classes = $\{i/i = 1, ..., n\}$) is trained separately using images of c_i . Figure 2 presents the flowchart of the first step.



Figure 2. Flowchart of the first step.

- In the second step, images intended to be used for SCNN training are generated. For each class, original images, their respective reconstructed versions are serve as positive examples as well as randomly selected images from other classes (i.e., other than the concerned class) and their respective reconstructed versions serve as negative instances, are prepared. Figure 3 presents the flowchart of the second step.
- In the third step, Siamese is trained using positive and negative instances prepared in the previous step. (Where the size of negative and positive examples that are fed to train Siamese is equal). Figure 4 presents the flowchart of the third step.

For a new probe p to be classified, a confidence score is generated for each class c_i by feeding p to $CAE_i + SCNN_i$. In particular, p is passed by each CAE_i to generate its reconstructed version p'. It is expected that once p and p' are passed. By the actual class (i.e., the class to which p belongs), a high similarity score will be yielded by the *SCNN*. On the contrary, passing by other classes yields low similarity scores that are near to zero. A p is assigned to the class having obtained the maximum confidence score.



Figure 3. Flowchart of the second step.



Figure 4. Flowchart of the third step.

• To speed up computations and reduce more the problem of inter species, we have opted for a hierarchical classification fashion by reducing the classification space based on clustering images from different classes using K-means [31]. For a test image, it is first mapped to the most appropriate cluster then matched solely with classes that belong to the selected cluster using our novel classifier (CAE base on Siamese as a loss). Figure 1 presents the two stages that the image feeds into our system.

3.2. Network Architecture and Loss Function

In our work, we have dealt with a convolutional auto-encoder that has been successfully applied to the computer vision domain. The convolutional auto-encoder is a subset of convolutional neural networks. They are similar, but the difference between them is that the weights in the CAE are shared among the inputs, preserving the spatial locality. Due to the use of CNN's integrated properties, some specific layers, such as convolutional, pooling, and so on, aid in feature extraction. Each convolution operation represents a filter that learns how to extract a specific plant feature by using filters. Following the convolutional operation, a pooling layer is usually included. The pooling layer reduces the input data's dimensionality.

In our work, for each class c_i we have designed a convolutional auto-encoder CAE_i , images within the class c_i are first fed to the encoder to generate the latent representation $(I_j \rightarrow R)$ through a series of convolutional and max-pooling layers, then second to the decoder with a series of de-convolutions and up-pooling layers to reconstruct the code $(R \rightarrow I'_i)$.

3.2.1. Convolutional Auto-Encoder Architecture

It consists of four convolutional blocks: convolutional layer with 8 filters with 10×10 , Relu activation function, and max pooling layer; convolutional layer with 16 filters with 5×5 , Relu activation function, and max pooling layer; convolutional layer with 32 filters with 3×3 , Relu activation function, and max pooling layer; convolutional layer with 64 filters with 3×3 , Relu activation function, and max pooling layer; and finally the Dense layer. Since we have to deal with CAE as a binary classifier (one class learning), the binary cross-entropy is used as a loss function (Equation (1)):

$$BCE(y, p) = -y \cdot \log(p) - (y - 1) \cdot \log(1 - p)$$
(1)

where *y* is the original image that we present to the AE and *p* is the reconstructed image obtained by the AE. For the decoder we have used de-convolutional and un-pooling layer it performs the inverse operation of the convolution layer and pooling layer. The network has symmetric architecture, with the same number of layers and feature maps generated in each layer in both parts.

Figure 5 gives an overview of CAE architecture. As it can be seen, this CAE is trained for class a which it could reconstruct images from this class well, on the contrary images from other classes (eg; b) can't be well reconstructed since it is not trained for.



Figure 5. CAE architecture.

3.2.2. Siamese Convolutional Neural Network Architecture, Loss Function

A Siamese neural network consists of two or more sub-networks that accept different inputs but are linked at the top by an energy function. A Siamese CNN consists of two symmetrical CNN neural networks both sharing the same weights and architecture. The objective of the Siamese network is to learn whether two input values are similar or dissimilar. Each CNN receives an input image, which is then processed through a series of convolutional and max-pooling layers. The last volume containing the extracted features is flattened into a 1D vector of features. A connected function will be used to connect the two vectors extracted by the convolutional neural network.

In our work on the top of CAE, we integrate a Siamese, which consists of two CNN as sub-networks of two convolutional blocks: convolutional layer with 32 filters with 10×10 , Relu activation function, and max-pooling layer; and convolutional layer with 64 filters

with 7 \times 7, Relu activation function, and Max pooling layer. The units of this convolutional layer are flattened into a single vector using global average pooling. This vector is then connected to a fully-connected layer (FCN) with 4096 neurons, a Relu activation function, and a soft-max layer. In order to merge the obtained two vectors *L*1 is used, from [32] the *L*1 distance since it is better than the *L*2 distance for Siamese based on CNN. A final dense layer (fully connected layer) with a sigmoid activation merges all the values into a single vector and produces the similarity or dissimilarity response. The dense layer computes a weighted sum of the vector's values (Equation (2)).

$$y = \sum_{i=1}^{N} (W(i) \times X(i)) + b$$
(2)

where $w_i/(i = 1, ..., n)$ represent the weights of the synapses of the dense layer; X_i (i = 1, ..., n) represent the elements of the merged vector achieved from l_1 (Equation (3)). distance.

$$l_1 = (|X1_1 - X2_1|, |X1_2 - X2_2|, \dots, |X1_n - X2_n|)$$
(3)

where X1 and X2 are the two vectors obtained from the two CNN; *n* represent the number of elements in each vector.

Then it adds a bias value to it and applies the sigmoid (σ) function (Equation (4)) to this value. The outcomes are in the interval of [0, 1]. The cross-entropy (Equation (1)) is exploited for training the network.

$$\sigma(y) = \frac{e_y}{e_y} + 1 \tag{4}$$

Figure 6 gives an overview of SCNN architecture. As it can be seen, this SCNN is trained on negative examples (images from class b and its reconstruction from CAE Figure 3) and positive examples (images from class a and its reconstruction from CAE). Siamese is trained with the output of y = 1 for positive examples and y = 0 for negatives.



Figure 6. Siamese architecture.

The integration of the two architectures presents our novel classifier that is based on CAE Using Siamese neural network as an alternative loss. Figure 7 present the architecture of our novel classifier.



Figure 7. Novel classifier (CAE based on Siamese as a loss) architecture.

3.3. Rationality of the Proposed Method

The main novelty in this work lies in considering a one-class learning strategy, where CAE and SCNN are integrated into a unified framework. Indeed, it is well-known that auto-encoder is an unsupervised network that is mainly used for dimensionality reduction and features learning. However, in our case, we extend it to a supervised scenario by considering loss values produced by each class. In particular, an auto-encoder is trained for each class separately, and the loss value is considered as an indicator of the class to which a test sample belongs. If a test sample is reconstructed for a class A with loss equals to x, and for class B with a loss equals to y, and x < y, then we can assign the sample to class A. Nevertheless, the CAE trained on complicated leaf images is eligible to punctually reconstruct those of relatively easier images from other classes. To overcome this problem, we propose using the Siamese neural network (Siamese for each class) as an intelligent loss metric on top of each CAE to alleviate the shortcomings of conventional loss. For the sake of illustration, the Siamese network of class 1 is trained using negative and positive examples, where positive is the original leaf image from class 1 and its reconstructed image, and the negative one is an image from other classes and its reconstructed one by the CAE of class 1. For a test image that is from class 1, as SCNN1 is trained on maximizing confidence score for original images from class 1 and their respective reconstructed versions (minimize the score for images from other classes, respectively), passing by CAE1 (+ Siamese 1) will almost produce a high similarity score. For another test sample from a class different than class 1, CAE1 (+ Siamese 1) will produce a low similarity score, as SCNN1 is trained to do so. As for testing strategy, some relevant works [13,14] have considered a hierarchical classification procedure, wherein classification is firstly performed for the coarse classes, and then passed to the subsequent levels in the hierarchy (i.e., genus, species). However, this procedure requires pre-knowledge of the dataset being classified, which is predominantly not possible. Thus, using those methods is limited to persons with knowledge of this field (i.e., experts). In this work, however, we adopt a hierarchical classification scheme that doesn't require this pre-knowledge, making it feasible to employ our method by non-experts. We cluster the whole dataset using K-means, such that each class falls exactly in one cluster. For a test image, instead of performing matching with all classes, it is mapped to the most appropriate cluster then matched only with classes that fall under the detected cluster. This permits a reduction in the classification space, and thus speeds up the recognition process.

4. Experiment

This section is devoted to evaluating our proposed approach under different conditions. In addition, we conduct a comparative evaluation against other recent works in order to prove the effectiveness of our method.

4.1. Dataset

ImageCLEF2012 leaf dataset: this dataset was created as part of the Pl@ntNet project. Images are collected from Western European regions. ImageCLEF2012 is one of the most challenging datasets due to its richness in terms of leaf categories (compound and simple realistic), species, variability, and similarity between species. As well as, differences on the acquisition level in terms of period, location and person. ImageCLEF2012 contains three types of images which are: scan, scan-like and photograph. The scan images have a white background, scan-like are images with minimal shadowing, and the photograph images are captured in nature with an uncontrolled manner. In our experiments, we have considered the scanned images, which represent 57% of the total database. There are 6630 images in this subset; 4870 in the training set are from 115 species and 1760 in the test set are from 110 species. Each class contains between 2 and 249 images. Figure 8 depicts representative samples from the ImageCLEF2012 scan dataset.



Figure 8. Representative samples from ImageCLEF2012 scan dataset.

4.2. Results

In this section, we will provide the results of our experiments. In all of the experiments, we have used the same protocol of the ImageCLEF2012 scan dataset. Approximately 4870 images are for training and 1760 images are for testing. From the training set, we have set 80% for training and 20% for validation.

4.2.1. Experiment 1: Measuring the Processing Time

As previously explained, the proposed model is made up of two components namely unsupervised convolutional auto-encoder (CAE) and the Siamese convolutional neural network that replaces the conventional loss function. The target of this experiment is to demonstrate that our model can converge rapidly, and Siamese doesn't require a high processing time. The Siamese curve of train and validation accuracy of class one of the first category are presented in Figure 9.

As can be seen from the two curves, our model can converge very quickly in more than 95% with only a few epochs. For example, at three and five epochs, the model achieved 93% and 94% for validation, respectively, and 96% and 97% for training, respectively. This proves that our system does not require a high processing cost and can learn to predict in a very short period of time.

To demonstrate that SCNN has improved our model and that it converges faster than a model based solely on CAE. The training and validation loss of CAE and Siamese are depicted in Figure 10. As can be seen in the first epoch, the CAE's loss for training is 34% and 30% for validation that is higher than the loss of SCNN which is only 21% for training and for 22% validation, two epochs after the SCNN has a training loss of 0.8% and a validation loss of 10% that is lower compared to 18% and 16% of the CAE's training and validation loss. From the curves, SCNN proves that it has improved our model and doesn't require a high processing time.



Figure 9. Training and validation accuracy curves.



Figure 10. Siamese and CAE loss training and validation curves: (a) Siamese and (b) CAE.

4.2.2. Experiment 2: Confusion Matrix

Regardless of the high performance that yields our method, it seems to occasionally confuse some leaf species. To discover the reason behind this confusion, we have to analyze the results in more detail. To this end, we have generated the confusion matrix of one category that contains 17 classes. The obtained confusion matrix are presented in Figure 11.

As it is illustrated in Figure 11, the confusion matrix provides more details about the evaluation outcomes. For instance, we can see that misclassification occurs among the species (21/41), (57/71), and (122/21) respectively. By taking a closer look at leaf images belonging to these species, we found out that some of their samples look visually identical in terms of color, veins and shape. Figure 12. Shows representative samples of a high color/geometric symmetry between leaves belonging to different species.

1	14	0	1	1	0	0	4	4	4	0	0	0	0	2	0	3	0
2	0	3	0	1	1	0	5	2	0	0	0	0	0	1	0	0	0
3	0	0	0	0	0	0	0	8	0	0	0	1	0	0	0	0	0
4	0	0	0	7	1	0	0	0	0	0	0	0	1	0	0	0	0
5	1	0	0	1	4	2	0	4	0	0	2	1	0	0	1	1	0
6	0	0	0	0	0	4	0	0	0	0	0	0	0	0	2	0	0
7	0	2	0	0	0	0	6	3	1	0	0	0	0	0	0	0	0
8	1	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0
9	0	2	0	1	1	0	2	2	1	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	1	0	19	0	0	0	0	0	0	0
11	0	0	2	1	0	2	1	0	0	0	3	0	0	0	2	1	0
12	0	0	0	0	0	0	0	1	0	0	0	8	0	0	2	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	7	0	3	0	0
14	1	0	0	0	0	1	0	0	1	0	0	0	0	10	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0
16	0	0	1	13	1	0	1	0	0	0	0	0	2	0	0	1	1
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17

Figure 11. Confusion matrix.



Figure 12. Representative samples of a high color/geometric correspondence between leaves belonging to different species.

4.2.3. Experiment 3: Comparison with State of the Art

In this sub-section, we aim to compare the proposed approach with other recent and relevant works on leaf classification. To do so, we opted for the ImageCLEF2012 dataset a configuration as in [17]. This subset contains 4870 leaf images for training and 1760 leaf images for testing. The performance evaluation standard employed was the same as [17]. Our comparison included a number of studies that were interested in leaf species and in the ImageCLEF2012 dataset. Figure 13 presents curves that give the results obtained by our novel method and the works involved.

Since we have dealt with categories and in order to get the final curve of the proposed method, we have first obtained the top ten of each category then, the sum of them represent our final curve. From the results, we can observe that our novel method (NM) achieves the best recognition performance among all of the competing methods in only the top ten. The recognition rate of our novel method is much higher compared to the deep learning methods VGG16 and Alex-Net [16], and the recognition accuracy of our method is higher than that of the two networks by 20% and 13%, respectively in just the top one. For the handcrafted methods, our system has also yielded very good results compared to that of the MDM [14], MTCD [4], and triangle-distance representation [16] methods; our system is higher than the mentioned methods by 22%, 21%, and 10%, respectively, when only one

candidate result is considered. These results indicate that the proposed approach using hierarchical methods based on one class learning techniques can distinguish different plant species very well. Furthermore, our novel system is well suited for large-scale images.



Figure 13. Recognition results on scan category of ImageCLEF2012 leaf Dataset.

5. Conclusions

An accurate hierarchical automatic recognition system based on a novel one-class learning classifier has been presented in this paper. In contrast to the conventional exploitation of CAE and Siamese in our case, we have proposed to extend CAE for novel supervised scenario by considering it as a one-class learning classifier in which a CAE is trained for each class and a Siamese is integrated as an alternative to the conventional loss of CAE. For each class, after training the CAE to reconstruct images from this class and to reconstruct images from other classes, Siamese is trained to distinguish the similarity and dissimilarity between the reconstructed leaf images from the trained class and the reconstructed images from the remaining classes. In contrast to the related hierarchical classification schemes, which require pre-knowledge of the dataset being recognized, our scheme consists of clustering the entire dataset to gather similar classes together. This strategy is simple, effective, and doesn't require experts or botanists. The performance of our system has been evaluated on well-known leaf datasets, namely ImageCLEF2012, and the results demonstrate that our approach exceeds existing state-of-the-art methods. A hierarchical representation has reduced the complexity of the process of classification and reduced inter-species problems. Furthermore, our novel one class learning classifier has outperformed the results of our system and the proposed intelligent loss; Siamese has exceeded the results of the CAE. Our perspective in future work will be driven by exploiting more objects that present hierarchical organization, and we expect to get reliable results from this approach.

Author Contributions: Conceptualization, L.H. and M.L.K.; methodology, L.H., M.L.K. and A.B.; software, L.H. and A.B.; validation, A.B.; writing—original draft, L.H.; writing—review and editing, O.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Chaki, J.; Parekh, R. Plant leaf recognition using shape based features and neural network classifiers. *Inter. J. Adv. Comput. Sc. Appl.* 2011, 2, 10. [CrossRef]
- 2. Aiadi, O.; Kherfi, M.L. A new method for automatic date fruit classification. Int. J. Comput. Vis. Robot. 2017, 6, 692–711. [CrossRef]
- Aiadi, O.; Kherfi, M.L.; Khaldi, B. Automatic Date Fruit Recognition Using Outlier Detection Techniques and Gaussian Mixture Models. *ELCVIA* 2019, 1, 52–75. [CrossRef]
- 4. Yang, C.; Wei, H.; Yu, Q. Multiscale Triangular Centroid Distance for Shape-Based Plant Leaf Recognition. In Proceedings of the European Conference on Artificial Intelligence, The Hague, The Netherlands, 29 August–2 September 2016; pp. 269–276.
- Hamrouni, L.; Bensaci, R.; Kherfi, M.L.; Khaldi, B.; Aiadi, O. Automatic recognition of plant leaves using parallel combination of classifiers. In Proceedings of the IFIP International Conference on Computational Intelligence and Its Applications, Oran, Algeria, 8–10 May 2018; pp. 597–606.
- Casanova, D.; de Mesquita Sá Junior, J.J.; Bruno, O.M. Plant leaf identification using Gabor wavelets. *Inter. J. Imaging Syst. Tech.* 2009, *3*, 236–243. [CrossRef]
- 7. Larese, M.G.; Namías, R.; Craviotto, R.M.; Arango, M.R.; Gallo, C.; Granitto, P.M. Automatic classification of legumes using leaf vein image features. *Pattern Recognit.* **2014**, *1*, 158–168. [CrossRef]
- 8. Kumar, T.P.; Reddy, M.V.P.; Bora, P.K. Leaf identification using shape and texture features. In *Proceedings of the International Conference on Computer Vision and Image Processing*; Springer: Singapore, 2017; pp. 531–541.
- 9. Zhang, H.; Tao, X. Leaf image recognition based on wavelet and fractal dimension. J. Comput. Syst. 2015, 1, 141–148.
- 10. Zhang, S.; Pan, X.; Cui, Y.; Zhao, X.; Liu, L. Learning affective video features for facial expression recognition via hybrid deep learning. *IEEE Access* **2019**, *7*, 32297–32304. [CrossRef]
- 11. Parekh, V.S.; Jacobs, M.A. Deep learning and radiomics in precision medicine. *Expert Rev. Precis. Med. Drug Dev.* **2019**, *2*, 59–72. [CrossRef]
- 12. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. Comput. Electron. Agric. 2018, 147, 70–90. [CrossRef]
- Araújo, V.M.; Britto, A.S.; Brun, A.L.; Koerich, A.L.; Oliveira, L.E. Fine-grained hierarchical classification of plant leaf images using fusion of deep models. In Proceedings of the 2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI), Volos, Greece, 5–7 November 2018; pp. 1–5.
- 14. Araujo, V.M.; Britto, A.S., Jr.; Oliveira, L.E.; Koerich, A.L. Two-View Fine-grained Classification of Plant Species. *arXiv* 2020, arXiv:2005.09110.
- 15. Zhang, S.; Huang, W.; Huang, Y.A.; Zhang, C. Plant Species Recognition Methods using Leaf Image: Overview. *Neurocomputing* **2020**, *408*, 246–272. [CrossRef]
- 16. Hu, R.; Jia, W.; Ling, H.; Huang, D. Multiscale distance matrix for fast plant leaf recognition. *IEEE. Trans. Image Proc.* **2012**, *11*, 4667–4672.
- 17. Wang, B.; Brown, D.; Gao, Y.; La Salle, J. MARCH: Multiscale-arch-height description for mobile retrieval of leaf images. *Inform. Sci.* **2015**, *302*, 132–148. [CrossRef]
- 18. Yang, C.; Wei, H. Plant species recognition using triangle-distance representation. IEEE Access 2019, 7, 178108–178120. [CrossRef]
- 19. Ghasab, M.A.J.; Khamis, S.; Mohammad, F.; Fariman, H.J. Feature decision-making ant colony optimization system for an automated recognition of plant species. *Exp. Syst. Appl.* **2015**, *5*, 2361–2370. [CrossRef]
- 20. Kadir, A.; Nugroho, L.E.; Santosa, P.I. Experiments of Zernike moments for leaf identification. JATIT 2012, 1, 82–93.
- 21. Naresh, Y.G.; Nagendraswamy, H.S. Classification of medicinal plants: An approach using modified LBP with symbolic representation. *Neurocomputing* **2016**, *173*, 1789–1797. [CrossRef]
- 22. Ambarwari, A.; Adria, Q.J.; Herdiyeni, Y.; Hermadi, I. Plant species identification based on leaf venation features using SVM. *Telkomnika* 2020, 2, 726–732. [CrossRef]
- Yigit, E.; Sabanci, K.; Toktas, A.; Kayabasi, A. A study on visual features of leaves in plant identification using artificial intelligence techniques. *Comput. Electron. Agric.* 2019, 156, 369–377. [CrossRef]
- 24. Mahajan, S.; Raina, A.; Gao, X.Z.; Kant Pandit, A. Plant Recognition Using Morphological Feature Extraction and Transfer Learning over SVM and AdaBoost. *Symmetry* **2021**, *2*, 356. [CrossRef]
- 25. Barré, P.; Stöver, B.C.; Müller, K.F.; Steinhage, V. LeafNet: A computer vision system for automatic plant species identification. *Ecol. Inform.* 2017, 40, 50–56. [CrossRef]
- 26. Hu, J.; Chen, Z.; Yang, M.; Zhang, R.; Cui, Y. A multiscale fusion convolutional neural network for plant leaf recognition. *IEEE Signal Process. Lett.* **2018**, *6*, 853–857. [CrossRef]
- 27. Lee, S.H.; Chan, C.S.; Mayo, S.J.; Remagnino, P. How deep learning extracts and learns leaf features for plant classification. *Pattern Recog.* **2017**, *71*, 1–13. [CrossRef]
- 28. Ghazi, M.M.; Yanikoglu, B.; Aptoula, E. Plant identification using deep neural networks via optimization of transfer learning parameters. *Neurocomputing* **2017**, 235, 228–235. [CrossRef]

- 29. Kumar, D.; Verma, C. Automatic Leaf Species Recognition Using Deep Neural Network. In *Evolving Technologies for Computing, Communication and Smart World*; Springer: Singapore, 2021; pp. 13–22.
- 30. Figueroa-Mata, G.; Mata-Montero, E. Using a convolutional Siamese network for image-based plant species identification with small datasets. *Biomimetics* 2020, *1*, 8. [CrossRef]
- Allaoui, M.; Kherfi, M.L.; Cheriet, A. Considerably Improving Clustering Algorithms Using UMAP Dimensionality Reduction Technique: A Comparative Study. In Proceedings of the International Conference on Image and Signal Processing, Marrakech, Morocco, 4–6 June 2020; pp. 317–325.
- Shih, C.H.; Yan, B.C.; Liu, S.H.; Chen, B. Investigating Siamese LSTM networks for text categorization. In Proceedings of the Pacific Signal and Information Processing Association Annual Summit and Conference, Kuala Lumpur, Malaysia, 12–15 December 2017; pp. 641–646.