

Multi-Organ Plant Identification With Multi-Column Deep Convolutional Neural Networks

Anfeng He, Xinmei Tian

CAS Key Laboratory of Technology in Geo-spatial Information Processing and Application System

University of Science and Technology of China, Hefei Anhui, China

Email: heanfeng@mail.ustc.edu.cn, xinmei@ustc.edu.cn

Abstract—Automatically identifying plants from images is a hot research topic due to its importance in production and science popularization. This process attempts to automatically identify the name of a plant with a known taxon from a given image. The majority of existing studies on automatic plant identification focus on identifying plants with a single organ, such as flower, leaf, or fruits. Plant identification using a single organ is not sufficiently reliable because different plants many have similar organs. To overcome this problem, this paper is devoted to automatically identifying plants by combining multiple organs of plants. Specifically, we propose a multi-column deep convolutional neural networks (MDCNN) model to combine multiple organs for efficient plant identification. Extensive experiments demonstrate the effectiveness of our model, and the plant identification performance is greatly improved.

I. INTRODUCTION

Plant identification consists of assigning a specimen plant with a known taxon. This process has evolved over hundreds of years but is a tedious task even for professional botanists. As shown in Fig. 1, it is not an easy task to identify the plants in these images. To narrow the botanical taxonomic gap, it is meaningful to automatically identify images of plants. Automatic plant identification has a wide variety of applications. For example, useful smartphone apps can be developed to identify photos of plants using an automatic plant identification algorithm for science popularization [1][2]. It can also help professional botanists to identify plants in the field during scientific research.

Automatic plant identification has attracted increasingly more attention in recent years with the rapid development of computer vision. Many research works have been conducted on automatically identifying plants [3][4][1]. Despite the progress achieved in this field, the performance of these works is far from satisfactory due to the challenge of this task and lack of robustness to easily confusable plants.

In early works, researchers primarily applied local features for image representation. For example, Szűcs et al. [3] used the SIFT (scale-invariant feature transform) [5] as the feature extractor on the dense grid and applied PCA (principal component analysis) [6] to reduce the dimension of the feature vector from 128 to 80. Then, the BoW (bag-of-words) [7] and GMM (Gaussian mixture model)[8] models were applied to describe the image. For the classification algorithm, the C-SVC with the RBF kernel was applied. However, the performance of

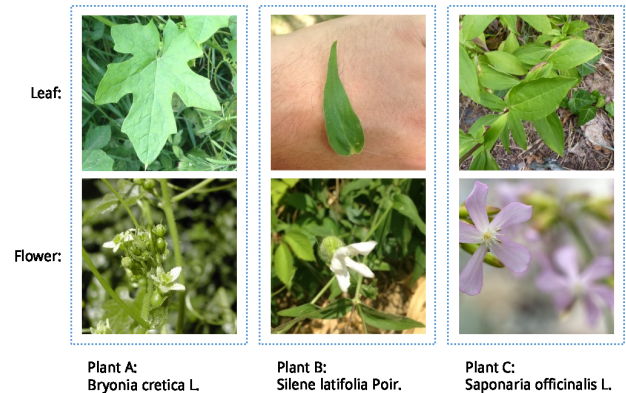


Fig. 1. The leaf and flower organs of three plants. Plant A and plant B have similar flowers but dissimilar leaves; plant B and plant C have dissimilar flowers but similar leaves.

these types of methods is unsatisfactory because of the limited discriminative description ability of low-level features. Kumar et al. [1] extracted curvature-based shape features from leaf images. Then, the nearest neighbor method was used for identification. Although this method achieved better performance, this solution heavily relies on the specifically selected set of handcraft features. Therefore, it may suffer from over-fitting and dataset bias problems [9][4].

Recently, DCNNs (deep convolutional neural networks) have been applied and achieved state-of-the-art performance in many image identification applications, such as ImageNet classification [10], hand-writing number classification [11], and face recognition [12]. Lee et al. [4] proposed training deep convolution neural networks to identify plants. Moreover, de-convolutional networks were employed to visualize the learning features for a better visual understanding of which features are important during leaf identification. Although this method outperforms other traditional methods, note that the dataset that they used is quite small and contains only 44 different plant species [4]. In addition, the current methods for plant identification using DCNNs were primarily focused on identifying a specimen plant from a single organ, e.g., flowers, leaves, or fruits, independently. The performance of plant identification using a single organ is unsatisfactory because a single organ cannot provide sufficient information for identification in some cases. As shown in Fig. 1, the

flowers of plant A and plant B are very similar, and thus, discriminating plant A from plant B by only using flowers is difficult. In other words, identifying plants using only a single organ may suffer from a noise problem; therefore, easily confused plants are likely to be misclassified. If we take both flower and leaf organs into consideration, plant A and plant B can be correctly identified.

In this paper, we apply the multi-column DCNN model to combine multiple organs for plant identification, to input more information to the classifier and to obtain more robust performance. Our multi-column DCNN model is flexible for any number of input organs. This flexibility means that regardless of the number of organs, the model can cope with it well.

The remainder of this paper is organized as follows. In Section II, we introduce DCNN and our proposed model in detail. The experimental results are presented in Section III. Finally, conclusions are drawn in Section IV.

II. MULTI-COLUMN DCNN FOR MULTI-ORGAN PLANT IDENTIFICATION

The multi-organ plant identification system contains two key stages. The first is that we should train separate sub-models for each type of organ. The second is to combine all single organ's models together efficiently. The entire structure of this system is shown in Fig. 2. Each row in the left part of this figure is an individual sub-model for a single organ. All seven models are combined together to train the integrated model. When testing, multiple organs of an unknown plant, rather than only one organ, are fed into the multi-column DCNN. Leveraging this powerful combination, the performance of plant identification is greatly improved.

A. Single-Organ Model

The identification performance from a single organ has a direct impact on the identification performance of multiple organs. Deep convolution neural networks have been proven to be an effective model for image classification problems [10]. Inspired by its powerful ability to describe image content well [13], we apply this model for plant identification in this paper.

CNN are feed-forward neural networks. It was first applied in the hand-written character recognition problem by LeCun et al. [14]. Deep convolutional neural networks have been developed in recent years with the rapid development of parallel computing. DCNNs are an efficient recognition method that have attracted widespread attention. The original image can be directly fed into the model. Therefore, this network can avoid the complicated pre-processing of the image. Convolutional neural networks make use of the spatial relationship to reduce the number of parameters. Compared to other feed-forward neural networks, convolutional neural networks require fewer parameters. However, many parameters still need to be trained for deep convolutional neural networks. Therefore, it is still challenging to train such a network with limited computing resources. Fortunately, the field of deep learning has achieved numerous significant improvements in recent years. For example, the dropout method proposed by Hinton et al. effectively

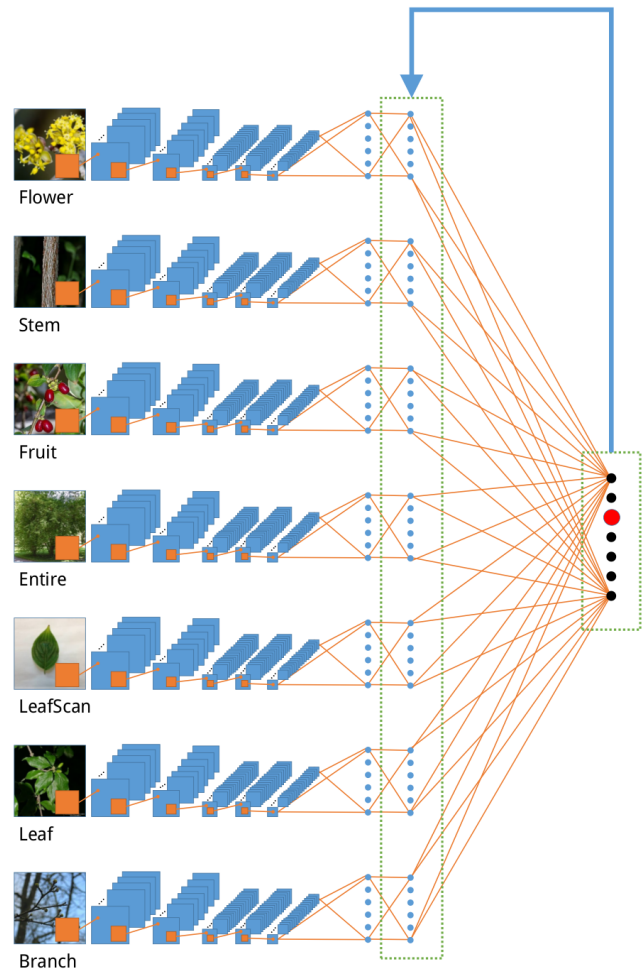


Fig. 2. The structure of the proposed multi-column deep convolution neural networks. Single-column networks are first trained. The input of each column is one type of specific organ. Seven organs are applied here, namely, flower, stem, fruit, entire, leafscan, leaf, and branch. The single-column networks are fine-tuned from AlexNet [10]. All models are combined together at the FC8 layer.

avoids the phenomenon of over-fitting by preventing the co-adaptation of feature detectors [15]. Furthermore, the rectified linear units (ReLU) achieve better performance than other active functions such as tanh [10][16]. Most importantly, the GPU implementation of deep convolution neural networks makes the training time as low as approximately 2.5 ms per image [17]. Due to these aforementioned methods, it is possible to efficiently train the deep convolutional neural networks. Based on these methods, deep convolutional neural networks with millions of parameters, proposed by Krizhevsky et al., were applied on the ImageNet dataset [10]. This network contains five convolutional layers and three fully connected layers. At the end of the network is the softmax layer, which outputs the probabilities of 1000 classes in the ImageNet dataset.

The deep convolution neural networks in this paper are

implemented as in [10]. We fine-tune the network from a pre-trained model for all organs of plants. The pre-trained models are trained on a large-scale image dataset [18], which contains approximately 1.2 million images of 1000 categories. With these abundant images, the pre-trained model can describe images well. Then, the pre-trained model is fine-tuned for plant identification of each organ. As shown in Fig. 2, there are seven organs used in the final model. Therefore, there are seven models that are fine-tuned for each of the seven organs.

B. Multiple Organ Combination via Multi-Column DCNN

As shown in Fig. 1, for each taxon of plants, information provided by a single organ is limited. Representing a taxon of a plant through multi-organ inputs can somewhat alleviate this problem. Each input of the model represents one type of organ of this taxon. Input with multiple organs can allow the model to obtain more information about this taxon of plants for identification.

Inspired by [19], [20] and [21], a novel multi-column deep convolutional neural network is proposed to combine different inputs of organs together. The architecture of this model is presented in Fig. 2. As shown in this figure, the models for different organs are independent in convolutional layers and the first two fully connected layers. The inputs of each column are the different organs of the plant. For those organs that did not appear in a taxon, we set the input of those columns with zero. The different organs for the initialized input are decided manually. The architecture of each column is exactly the same as in [10]. The Fc8 layers of all columns are combined together with a fully connected layer. The output of the combined fully connected layer is a softmax layer that outputs the probabilities of each taxon.

However, after all sub-models are combined together, the number of parameters in this model is seven times larger than that in the single column. To implement this model and obtain better performance, the following tricks are applied in this model.

1) *The weights in Fc8 of the multi-column model is initialized from the weights of the single-column model:* The weights of the single-column model are fine-tuned well from AlexNet [10]. It is clearly better than randomly initialized weights.

2) *Probabilities of each column restrict the combined weights:* The combination of multiple organs is based on the assumption that most of the organs can provide output with a high confidence level. However, when more organs are fed into the model, we cannot ensure that most of the organs can be output with high-confidence results. It is possible that the high-confidence outputs will be lost in the majority of low-confidence outputs. Consequently, the low-confidence outputs should be restricted. The probability output of each column can represent not only the probabilities of each taxon but also the confidence level of this prediction. From this perspective, the confidence level of the i^{th} column \mathcal{CL}_i is simply defined by the maximum of the probability out \mathcal{P} of image in the i^{th} column \mathcal{I}_i . It can be written as follows:

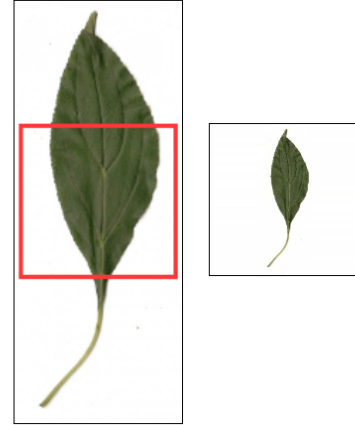


Fig. 3. The size of the input images of the deep convolutional neural networks is fixed. Two types of methods to normalize the size of images are shown. The method on the left is cropping the center square then resizing this patch to 256×256 . The method on the right is resizing the image to ensure that the length of the long edge is 256. Then, the image is placed in the center of 256×256 patch. The empty area of the image is filled by pixels whose values are the same as in the long edge.

$$\mathcal{CL}_i = \max(\mathcal{P}(\mathcal{I}_i)) \quad (1)$$

Prior to combining, the output of Fc7 in the i^{th} column whose confidence level \mathcal{CL}_i is less than a pre-defined threshold θ is multiplied by \mathcal{CL}_i :

$$O_i^{Fc7} = \begin{cases} O_i^{Fc7} & \text{if } \mathcal{CL}_i \geq \theta \\ O_i^{Fc7} \times \mathcal{CL}_i & \text{if } \mathcal{CL}_i < \theta \end{cases} \quad (2)$$

3) *Crop center image for organs except the LeafScan:* The size of the input images for deep convolutional neural networks is fixed. All images are cropped to 256×256 . However, as shown in Fig. 3, the crop of “LeafScan” images will lose most parts of the “LeafScan” images. To overcome this problem, rather than cropping the “LeafScan” images, we resize the image to ensure that the long edge equals 256. The board is filled with the same pixels as these on the long edge.

4) *Oversample image when identifying:* The input images are predicted from crops of the middle, top left, top right, bottom left and bottom right. The mirrored image is also cropped and predicted. The final result is the average of these ten predictions.

Using these tricks, the model can be flexible for various types of inputs. Therefore, this model can achieve better performance and is more robust.

III. EXPERIMENT

This section verifies the performance of our multi-column deep convolutional neural networks for multi-organ plant identification.

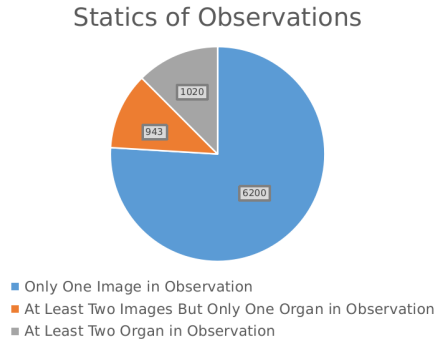


Fig. 4. Statistics of observations. The total number of observations is 8163. However, there are only 12.5% observations that have at least 2 organs.

A. Dataset

Our experiment is conducted on the plant task of the Life-CLEF2014 dataset:PlantCLEF [22]. This dataset is different from the previous plant identification datasets. Queries in this dataset are defined as plant observations rather than single images. The plant observations are a set of 1 to 5 images from the same individual plant observed by the same person on the same day. Some examples of observations are presented in Fig. 5. Each image in an observation belongs to a single view type (flower, fruit, branch, leaf, entire plant, stem, or leaf scan). This dataset contains 47815 images for training (6356 images of “Entire”, 3753 of “Fruit”, 1987 of “Branch”, 13164 of “Flower”, 3466 of “Stem”, 7754 of “Leaf”, and 11335 scans and scan-like pictures of leaf) and 8163 plant-observation-queries with 13146 images (731 images of “Branch”, 2058 of “Leaf”, 4559 of “Flower”, 2983 of “Entire”, 1184 of “Fruit”, 935 of “Stem” and 696 scans and scan-like pictures of leaf) for testing. Fig. 4 shows the statistics of observations. As shown in this figure, the number of observations whose contents have at least two organs is quite limited. Therefore, we pair-wise combine the organs rather than combining all organs together.

B. Performance Metrics

In addition to the basic top-1 and top-5 identification accuracy metrics, to compare our model with other models using this dataset, a new score metric is used in the experiments. Each plant observation test will be given a score $s_{u,p}$ for the p^{th} plant observed by the u^{th} user. The score $s_{u,p}$ is equal to the inverse of the rank of the correct species. An average score S will then be computed on all tested individual plants. Because the task wants to evaluate the ability of a system to provide correct answers to all users, the score is the mean of the average classification rate per author. Finally, the primary metric is defined as the following average classification score S_O :

$$S_O = \frac{1}{U} \sum_{u=1}^U \frac{1}{P_u} \sum_{p=1}^{P_u} s_{u,p} \quad (3)$$



Fig. 5. Some examples of observations.

where P_u is the number of individual plants observed by the u^{th} user and U is the number of users who have at least one image in the test data.

The score averaged over all images rather than over all observations is defined as follows:

$$S_I = \frac{1}{U} \sum_{u=1}^U \frac{1}{P_u} \sum_{p=1}^{P_u} \frac{1}{N_{u,p}} \sum_{n=1}^{N_{u,p}} s_{u,p,n} \quad (4)$$

where $N_{u,p}$ is the number of pictures taken from the p^{th} plant observed by the u^{th} user and $s_{u,p,n}$ is the score equal to the inverse of the rank of the correct species for the n^{th} picture taken from the p^{th} plant observed by the u^{th} user.

C. Comparison on Single-organ and Multi-organ Models

To evaluate how much progress is made by multi-organ input, we compare the results of multiple organs with the results of a single organ. These results are summarized in Table I and Table II.

The “MC” result is the multi-column deep convolutional neural networks proposed in this paper. The “AVGP” result is simply calculated as the mean of the query pair’s probabilities. Compared to the results of a single organ, the results show that irrespective of the type of method applied to combine the model, most of the results of multiple organs are far better than the results of a single organ. Moreover, the results of the multi-column deep convolutional neural networks proposed in this paper achieve the best performance. However, the results show that some combinations cause the performance to become worse. For example, the performance of the combination of

TABLE I
IDENTIFICATION ACCURACY OF SINGLE ORGAN

Organ	Top 1 (%)	Top 5 (%)
LeafScan (LS)	64.4	86.1
Flower (Fl)	61.0	81.0
Fruit (Fr)	35.6	59.0
Leaf (Le)	29.0	52.0
Stem (St)	29.7	44.5
Branch (Br)	16.8	35.8
Entire (En)	29.2	51.0

TABLE II
IDENTIFICATION ACCURACY OF TWO ORGANS

Accuracy (%)		Fl	Fr	Le	St	Br	En	
LS	MC	T1	79.3	76.7	71.4	66.3	56.9	51.6
		T5	91.7	90.6	90.1	85.0	81.0	71.7
	AVGP	T1	73.4	65.9	64.3	61.7	56.0	46.5
		T5	89.4	88.2	85.9	81.9	79.3	68.8
Fl	MC	T1	—	75.0	74.2	64.2	64.1	73.7
		T5	—	91.1	90.5	82.7	84.0	90.8
	AVGP	T1	—	66.2	64.7	60.2	59.2	66.4
		T5	—	85.5	83.5	79.4	79.2	84.1
Fr	MC	T1	—	—	64.3	51.4	50.0	50.8
		T5	—	—	84.5	73.9	74.1	74.3
	AVGP	T1	—	—	50.9	45.5	41.2	40.6
		T5	—	—	76.3	68.0	64.8	64.1
Le	MC	T1	—	—	—	51.0	39.8	46.7
		T5	—	—	—	73.7	67.5	70.8
	AVGP	T1	—	—	—	43.6	34.1	34.3
		T5	—	—	—	66.5	59.4	59.7
St	MC	T1	—	—	—	—	31.1	40.5
		T5	—	—	—	—	52.7	62.3
	AVGP	T1	—	—	—	—	29.1	35.9
		T5	—	—	—	—	46.3	57.3
Br	MC	T1	—	—	—	—	—	34.2
		T5	—	—	—	—	—	58.8
	AVGP	T1	—	—	—	—	—	28.5
		T5	—	—	—	—	—	50.7

“Entire” and “LeafScan” is lower than the performance of “LeafScan”. To explain this phenomenon, we counted the testing dataset. We found that the reason for this phenomenon was that there were only few taxa that had the organ “LeafScan”. Therefore, the identification accuracy does not benefit from the combination.

D. Comparison of Our Method with State-of-the-art

As previously mentioned, to compare our results with other state-of-the-art methods, we apply the score matrices on the top-20 results. Then, we compare the score to the best results in the PlantCLEF2014 challenge task.

1) *Compared to single organ:* For the results shown in Table III, the “FT-CNN” is our proposed method. The “SIFT+FV+CNN” is a model proposed by the best result in

TABLE III
IDENTIFICATION SCORE IN SINGLE ORGAN COMPARISON BETWEEN BASELINES AND OUR MODELS.

Organ	CNN [23]	SIFT+FV+CNN [23]	FT-CNN (ours)
LeafScan (LS)	0.278	0.640	0.641
Flower (Fl)	0.389	0.585	0.648
Fruit (Fr)	0.161	0.339	0.472
Leaf (Le)	0.103	0.318	0.329
Stem (St)	0.138	0.269	0.224
Branch (Br)	0.103	0.292	0.235
Entire (En)	0.193	0.333	0.394
Average	0.263	0.456	0.488

TABLE IV
IDENTIFICATION SCORE IN OBSERVATIONS COMPARISON BETWEEN BASELINES AND OUR MODELS.

Method	Score
CNN [23]	0.271
SIFT+FV+CNN [23]	0.471
FT-CNN+AVGP (ours)	0.503
FT-CNN+MC (ours)	0.506
FT-CNN+MC+RLCL (ours)	0.508

the PlantCLEF2014 task [23], which uses dense SIFT as the feature descriptor. Feature representation uses the GMM-based Fisher vector. A newly trained CNN feature is added to the feature vector. The “CNN” is the results of the newly trained CNN. The results show that our method with a single organ is better than the best result in the PlantCLEF2014 task.

2) *Comparison of observations:* Table IV presents a comparison of our and the best results in the PlantCLEF2014 task [23]. The “FT-CNN+AVGP” is the performance of the average probability of observations output. The “FT-CNN+MC” is the performance of the multi-column deep convolutional neural networks proposed by us. The “FT-CNN+MC+RLCL” is the performance of our model, and weights are added to each column according to the confidence level of the output. Although the growth rate is not very large due to the observation data with multiple organs being seldom accounted for, the result shows that our model with added weights is the best. As shown in Fig. 6, the result of leaf and flower of this taxon is wrong in top-1; however, when the input observation consists of these two images to our proposed model, the result is correct. This example demonstrates that plant identification with multiple organs is powerful for providing the correct identification even when both identifications of a single input are incorrect.

IV. CONCLUSION

In this paper, the limitations of deep convolutional neural networks in plant identification and the benefits of multi-column deep convolutional neural networks are analyzed. Subsequently, an effective automatic plant identification algorithm with multiple organs by using multi-column deep convolutional neural networks is proposed. The single model is fine-tuned

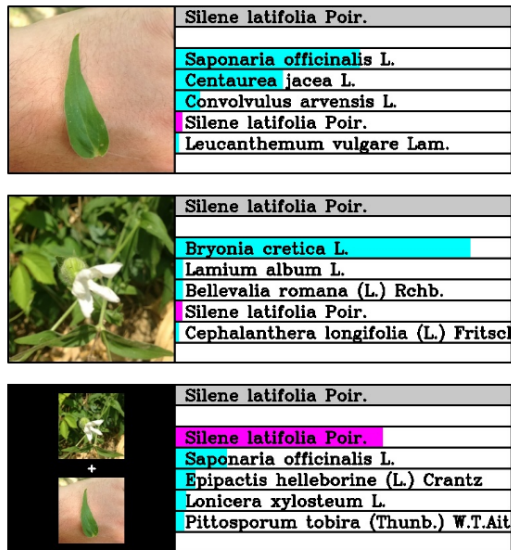


Fig. 6. Comparison of single-organ classification and multi-column classification. There are two single-organ tests in the first row and the second row. One multi-organ test is presented in the third row. The first label in each line is the ground truth of this image. The remainder of the labels are the five labels predicted with the highest probability. The bar is purple if the label is the same as the ground truth.

from the pre-trained model. The experiments conducted on the plant dataset that consists of user's observation queries in the real world have demonstrated the superiority of our proposed model.

In the future, we will continue exploring the automatic plant identification problem with different methods. The performance of some single organs is quite lower than that of other organs, which limits the overall performance of multi-organ identification. Therefore, it is necessary to explore useful methods to improve the performance of organs whose performance is low. Furthermore, AlexNet can be replaced by some more powerful models, e.g., GoogleNet [24] and VGG-Net [25]. We believe that this automatic plant identification system will bring more reliable applications.

ACKNOWLEDGMENT

This work is supported by the NSFC under the contracts No.61572451 and No.61201413, Youth Innovation Promotion Association CAS CX2100060016, Fok Ying Tung Education Foundation, the Specialized Research Fund for the Doctoral Program of Higher Education No.WJ2100060003, the Fundamental Research Funds for the Central Universities WK2100060011 and WK2100100021.

REFERENCES

- [1] N. Kumar, P. N. Belhumeur, A. Biswas, D. W. Jacobs, W. J. Kress, I. C. Lopez, and J. V. Soares, "Leafsnap: A computer vision system for automatic plant species identification," in *Computer Vision—ECCV 2012*. Springer, 2012, pp. 502–516.
- [2] A. Joly, H. Goëau, P. Bonnet, V. Bakic, J. Barbe, S. Selmi, I. Yahiaoui, J. Carré, E. Mousset, J.-F. Molino *et al.*, "Interactive plant identification based on social image data," *Ecological Informatics*, vol. 23, pp. 22–34, 2014.
- [3] G. Szűcs, D. Papp, and D. Lovas, "Viewpoints combined classification method in image-based plant identification task," 2014.
- [4] S. H. Lee, C. S. Chan, P. Wilkin, and P. Remagnino, "Deep-plant: Plant identification with convolutional neural networks," in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 452–456.
- [5] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [6] I. Jolliffe, *Principal component analysis*. Wiley Online Library.
- [7] F.-F. Li, R. Fergus, and A. Torralba, "Recognizing and learning object categories," *Tutorial at ICCV*, 2005.
- [8] D. Reynolds, "Gaussian mixture models," *Encyclopedia of Biometrics*, pp. 827–832, 2015.
- [9] A. Torralba and A. A. Efros, "Unbiased look at dataset bias," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 1521–1528.
- [10] 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.
- [11] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [12] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1701–1708.
- [13] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, "Decaf: A deep convolutional activation feature for generic visual recognition," in *Proceedings of The 31st International Conference on Machine Learning*, 2014, pp. 647–655.
- [14] B. B. Le Cun, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Handwritten digit recognition with a back-propagation network," in *Advances in neural information processing systems*. Cite-seer, 1990.
- [15] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *arXiv preprint arXiv:1207.0580*, 2012.
- [16] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 807–814.
- [17] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the ACM International Conference on Multimedia*. ACM, 2014, pp. 675–678.
- [18] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 248–255.
- [19] H. Goëau, P. Bonnet, J. Barbe, V. Bakic, A. Joly, J.-F. Molino, D. Barthelemy, and N. Boujemaa, "Multi-organ plant identification," in *Proceedings of the 1st ACM international workshop on Multimedia analysis for ecological data*. ACM, 2012, pp. 41–44.
- [20] J. Schmidhuber, "Multi-column deep neural networks for image classification," in *IEEE Conference on Computer Vision & Pattern Recognition*, 2012, pp. 3642–3649.
- [21] X. Lu, Z. Lin, H. Jin, J. Yang, and J. Z. Wang, "Rapid: Rating pictorial aesthetics using deep learning," in *Proceedings of the ACM International Conference on Multimedia*. ACM, 2014, pp. 457–466.
- [22] A. Joly, H. Müller, H. Goëau, H. Glotin, C. Spampinato, A. Rauber, P. Bonnet, W.-P. Vellinga, R. B. Fisher, and R. Planquè, "Lifeclef: Multimedia life species identification," in *EMR@ ICMR*, 2014, pp. 7–13.
- [23] Q. Chen, M. Abedini, R. Garnavi, and X. Liang, "Ibm research australia at lifeclef2014: Plant identification task," in *CLEF (Working Notes)*, 2014, pp. 693–704.
- [24] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.
- [25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.