Data Augmentation for Plant Classification

Pawara, Pornntiwa; Okafor, Emmanuel; Schomaker, Lambertus; Wiering, Marco

Published in:
Advanced Concepts for Intelligent Vision Systems (Acivs 2017)

Data Augmentation for Plant Classification

Pornntiwa Pawara, Emmanuel Okafor, Lambert Schomaker, and Marco Wiering

Institute of Artificial Intelligence and Cognitive Engineering (ALICE)
Nijenborgh 9, University of Groningen, The Netherlands
{p.pawara,e.okafor,l.r.b.schomaker,m.a.wiering}@rug.nl

Abstract. Data augmentation plays a crucial role in increasing the number of training images, which often aids to improve classification performances of deep learning techniques for computer vision problems. In this paper, we employ the deep learning framework and determine the effects of several data-augmentation (DA) techniques for plant classification problems. For this, we use two convolutional neural network (CNN) architectures, AlexNet and GoogleNet trained from scratch or using pre-trained weights. These CNN models are then trained and tested on both original and data-augmented image datasets for three plant classification problems: Folio, AgrilPlant, and the Swedish leaf dataset. We evaluate the utility of six individual DA techniques (rotation, blur, contrast, scaling, illumination, and projective transformation) and several combinations of these techniques, resulting in a total of 12 data-augmentation methods. The results show that the CNN methods with particular data-augmented datasets yield the highest accuracies, which also surpass previous results on the three datasets. Furthermore, the CNN models trained from scratch profit a lot from data augmentation, whereas the fine-tuned CNN models do not really profit from data augmentation. Finally, we observed that data-augmentation using combinations of rotation and different illuminations or different contrasts helped most for getting high performances with the scratch CNN models.

Keywords: Plant Classification, Data Augmentation, Deep Convolutional Neural Networks

1 Introduction

Plant classification using machine learning and computer vision algorithms is concerned with categorizing plant images into identifiable groups. This may help people to know for example the name of a tree they encounter based on a picture from a leaf of the tree. The classification problem can be challenging because of issues related to a high inter-class similarity, intra-class diversities, possible variations of complex backgrounds, and color and illumination variations within the image dataset. Previous studies have employed several supervised learning algorithms combined with hand-crafted features [6], [9], [17], [28] and global features [2] for investigating plant identification. An extension of the use of the

hand-crafted features is the combination of geometric-based features with a probabilistic neural network for classifying different classes of the Foliage dataset [7]. The recent advances in deep learning [5] have led to some big successes in several plant recognition studies [3], [4], [14]. The authors in [14] have investigated the use of the famous CNN architectures AlexNet [8] and GoogleNet [25] for plant classification. Moreover, the research in [4] considered the previous architectures and VGGNet [22] in their plant classification task. Generally, CNN architectures consist of many layers and have millions of parameters in the network [10]. Therefore, they need large datasets during the learning process.

Several works [4], [13], [20] have shown that increasing the number of images in the training set with data-augmentation (DA) techniques is useful to reduce overfitting and improve the overall performance of the CNN models. The fundamental idea is that the object of interest in an image will not change its class if the image is somewhat changed using a particular image-processing operation. Data augmentation can be performed in many ways, e.g. using translation, rotation, change in illumination, and color casting and processed in two stages: off-line and online [21]. Off-line augmentation involves an increase in the number of training images before the training starts, while the online stage increases the number of image appearances during the training process. The authors in [11] performed off-line augmentation by rescaling the training images into three different sizes and cropped them into smaller-sized images and combined this with horizontal flips for creating the augmented images during training. The leaf classification system in [23] employed three data-augmentation techniques: affine and perspective transformation, and rotation during the training stage. However, there has been little research to investigate the effects of many different single and combined data-augmentation methods such as combining different pose and illumination variants, in order to determine if this helps the CNNs to obtain significantly better performances.

Contributions: In this paper, we examine the effects of different data-augmentation techniques using two off-the-shelf CNN techniques: AlexNet and GoogleNet, which we train from scratch or using pre-trained weights. For this, we use three different image datasets of plants and we evaluate the CNNs on the original datasets, the datasets obtained using a single DA technique, and the datasets obtained using several combinations of DA techniques. Note that the DA techniques are only applied on the training data, therefore this results in 12 training set variants for the three plant recognition datasets. The results show that when the CNN methods are trained from scratch, the use of DA techniques is very effective to obtain higher performances. Especially combinations of the rotation and illumination DA techniques or rotation and contrast are most useful for the considered datasets. For the fine-tuned CNN models, the gains of DA techniques are much smaller, although they helped to get the best results, which are also better than previous results on the three plant datasets.

Paper Outline. The rest of the paper is organized as follows. Section 2 covers details of the three plant datasets used in this study and the different data-augmentation techniques. The CNN methods and experimental settings

are described in section 3. The results are shown and discussed in section 4. Finally, we draw a conclusion and recommend future work in section 5.

2 Datasets and Data-Augmentation Techniques

In this section, we describe the three plant datasets and the data-augmentation techniques which are used in the experiments. In Figure 1, we show some examples of images within the datasets.

2.1 Datasets

The Folio Dataset: Folio [16], a relatively small dataset, consists of 637 leaf images from 32 species. Each class contains approximately 20 images (three images are missing from the initial work of [16]). All images were taken under daylight on a plain background. The first classification system for this dataset used shape features and a color histogram with a k-nearest neighbor classifier [16] and reported an accuracy of 87.3%. The most recent study in [19] employed CNN techniques applied on the original images. The best CNN architecture obtained a high accuracy of 97.7%. We used the same train/validation/test splits as in [19] with a ratio of 70:10:20.

The AgrilPlant Dataset: The AgrilPlant dataset was presented in [19] and it consists of 3,000 plant images from 10 classes: apple, banana, grape, jackfruit, orange, papaya, persimmon, pineapple, sunflower, and tulip. Each class consists of 300 images. The AgrilPlant dataset faces some challenges due to the following reasons: 1) a dissimilarity of plants within the same class, for example, there are varieties of shape and color of tulips, or there are several colors of apples, 2) a similarity among some classes, for example, apple, orange, and persimmon images consist of similar shapes and colors, and 3) the complex backgrounds in most of the images. We adopted the same dataset splits as previously used in [19] with a ratio of 70:10:20 for train, validation, and testing sets, respectively.

The Swedish Dataset: The Swedish dataset [24] contains 1,125 plant leaf images on a plain background of 15 different Swedish tree species, with 75 images per class. The earlier research in [24] combined simple features such as moments, area and curvature and reported an accuracy of 82%. To the best of our knowledge, the study in [26] yielded the highest accuracy of 99.5%. This was achieved by combining shape, color, and Haralick features.

The authors in [1] proposed CNN methods with horizontal flip augmentation on this dataset and this obtained an accuracy of 99.1%. The challenge of classification on the Swedish dataset [15], [27], [29] is its high inter-species similarity among several classes. Our study used the same dataset splits as in [24] with randomly selecting 25 images per class for training and the rest for testing. Additionally, the training images were further dissected in the ratio 1:4 for validation and training sets, respectively.

Fig. 1. Some example pictures from the three datasets in which we show one image per class for some classes in the datasets. From the top row to the bottom row we can see example images from the Folio, AgrilPlant, and Swedish datasets.

2.2 Data Augmentation

In this subsection, we describe the six different data-augmentation techniques examined with the goal to increase the number of images within the training set for each of the datasets discussed in the previous subsection. The data-augmentation techniques we studied in this paper are:

Rotation: Our preliminary experiments were done on the AgrilPlant dataset. Using different rotational angles that exist between 8° and 90°, we observed that using a tilt of an image with angle 30° obtained good performances. This is the reason for the choice of using random image rotations with a rotation angle in [-30°, 30°], with empty space padded with white pixels.

Blur: The goal of the blur augmentation is to de-emphasize differences in adjacent pixel values. In this paper, the 2D Gaussian smoothing kernel is used. The kernel size is set to $2 \times (\lceil 2\sigma \rceil) + 1$, where $\lceil . \rceil$ is a ceiling function, and σ is the standard deviation of the Gaussian distribution which is randomly set between 2 and 8.

Scaling: The training images are rescaled to larger ones with a random factor between 2 and 8 times. Hence, when feeding the images into the CNNs, we crop the images from the up-scaled images and this corresponds to a small subpart of the image which may contain important features of the plants.

Contrast: We first convert images from an RGB color map to an HSV color map, then multiply the S and V components of the images by a random factor between 0.8 and 2. Finally, the images are converted back to the RGB color representation.

Illumination: The training images are adjusted by adding random values between 10 and 80 to the R, G and B channels.

Projective: The projective transformation changes the projective viewpoint of the observer. After transformation, straight lines still remain straight [23] but

it does not preserve parallelism, length, and angle. The projective transformation requires a 3×3 transformation matrix[1].

$$(x_j, y_j, 1) = (x_i, y_i, 1) \times \begin{pmatrix} cos(\theta) & sin(\theta) & t_1 \\ sin(\theta) & cos(\theta) & t_2 \\ 0 & 0 & 1 \end{pmatrix} \quad (1)$$

where $(x_i, y_i, 1)$ represents the coordinate before the projective transformation, $(x_j, y_j, 1)$ denotes the coordinate after the transformation, θ is the rotation angle of the image, and $[t_1 \ t_2]^T$ is the projection vector which is set to $[0.001 \ \ 0.001]^T$. The angle θ is randomly chosen from the interval $[1, 30]$.

The effects of all DA techniques on some example images of the AgrilPlant dataset are shown in Figure 2. In addition to the use of these single DA methods, we also consider several combinations of the earlier discussed methods to obtain more training images. Because testing all combinations is almost infeasible, we tested only combinations in which the rotation operator is part of the combined DA technique. This results in six possible combinations of DA methods which include: rotation+blur, rotation+contrast, rotation+scaling, rotation+illumination, rotation+projective, and rotation+contrast+illumination. Each single data-augmentation method adds eight adapted copies of the original images while the combination of two DA methods results in 16 different copies of the images. Lastly, the combination of three DA methods yields 24 times more training images. The total number of images present in each of the original and the DA image datasets are summarized in Table 1.

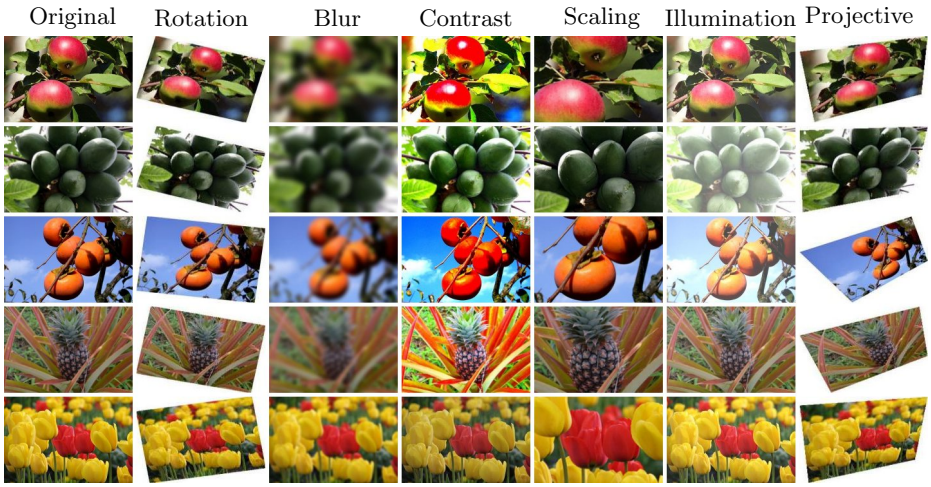

Fig. 2. Effects of data augmentation on some example images of the AgrilPlant dataset.

[1] https://www.graphicsmill.com/docs/gm5/Transformations.htm

Table 1. Summary of the number of training images in the (data-augmented) datasets.

Data set	Folio	AgrilPlant	Swedish
Original	445	2100	300
Individual DA	4,005	18,900	2,700
Combination of two DAs	7,565	35,700	5,100
Combination of three DAs	11,125	52,500	7,500

3 Deep Learning Architectures

3.1 CNN Methods

In our study, we employ two CNN architectures: AlexNet and GoogleNet for evaluating both original and several variants of data-augmented image datasets for the three plant recognition tasks.

AlexNet: The CNN architecture AlexNet [8] outperformed other computer vision methods during the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012. The network consists of five convolutional layers, three max pooling layers, two dropout layers, and three fully connected layers ending with a SoftMax classification layer. It uses the Rectified Linear Unit (ReLU) for the non-linear activation functions. In our study, we employed a customised version of AlexNet as proposed in [19], in which we reduced the number of hidden units in the last fully connected layers to 1024 neurons. We also consider two instances of the AlexNet architecture: using randomly initialized weights (scratch) and using pre-trained weights (fine-tuned). In the fine-tuned network, the pre-trained weights from ImageNet were used, after which we trained the whole architecture based on the errors for classifying the training images from the plant datasets.

GoogleNet: GoogleNet [25] is a deeper network, but has a much lower number of parameters (4 million parameters) compared to AlexNet (60 million parameters). This is a consequence of the inception module that vastly decreases the amount of trainable parameters in the network. More specifically, GoogleNet uses nine inception modules, four convolutional layers, four max-pooling layers, three average pooling layers, five fully connected layers and three SoftMax layers for the main and auxiliary classifiers in the network. Inspired by the network-in-network approach [12], the inception module uses a parallel combination of 1×1, 3×3, and 5×5 convolutions along with a pooling layer. A more detailed explanation and all relevant parameters of the GoogleNet architecture can be found in the original paper [25]. Similarly as with AlexNet, we evaluated both scratch and fine-tuned versions of the GoogleNet architecture.

3.2 Experimental Setup

We evaluate the deep CNN architectures with the different data-augmentation schemes for the three plant classification tasks. In the experiments, we employed 5-fold cross validation to evaluate the performances of the different methods. The resolution of the images is set to 256×256 pixels.

The AlexNet and GoogleNet hyper-parameters are set as follows: number of iterations: 20,000 for fine-tuned and 50,000 for the scratch version, step size: 10,000 and 25,000 for fine-tuned and scratch, respectively, train batch size: 20, validation batch size: 10, base learning: 0.001, momentum: 0.9, weight decay: 0.0005, and test interval: 10,000. Each dataset contains a different number of images, therefore we set different batch sizes for the different datasets as 7, 30 and 8 for Folio, AgrilPlant, and Swedish, respectively.

To summarize, we performed a total of 52 experiments on each dataset, which vary in the following settings: two choices of deep learning architecture (AlexNet and GoogleNet), two choices of training mechanism (fine-tuned or scratch), using the set of original images, and 12 datasets constructed with different data-augmentation techniques (as described in section 2.2).

4 Results

In this section, we report the test accuracies using the deep learning methods on the original and augmented datasets for the different plant recognition tasks. We report the top-1 accuracy and average the results over the five folds.

4.1 Folio Dataset Evaluation

Table 2 shows the plant classification accuracies with different DA techniques on the Folio dataset using AlexNet and GoogleNet with both scratch and fine-tuned models. The scratch AlexNet always profits from the different DA techniques on this dataset, whereas scratch GoogleNet also profits from most DA techniques, but in a lesser degree. Scratch AlexNet profits most from the combined effects of rotation and illumination, or combined effects of rotation, contrast, and illumination which led to a performance improvement of around 8.8% compared to using the original images. The best single DA technique for scratch AlexNet is the illumination operator, and blur is the DA technique that helps the least in getting higher performances. For scratch GoogleNet the best DA technique uses the scaling operation and this leads to 1.5% accuracy improvement compared to training on the original images. For the fine-tuned architectures, GoogleNet with the illumination DA technique obtains the highest accuracy. Because the fine-tuned models already perform very well with the original dataset, the improvements are much smaller in this case than when using the scratch CNN architectures.

When we compare our approaches to previous CNN experiments in [19], which did not consider flipping of the images, these new results show a significant improvement in the recognition performance. This shows that the effect of flipping is also very important for this dataset and that the offline DA techniques can help to obtain even higher performances.

Table 2. Recognition results (accuracy and standard deviation) using different DA schemes for the Folio dataset.

Augmentation methods	AlexNet		GoogleNet	
	Scratch	Fine-tuned	Scratch	Fine-tuned
Original (no flip) [19]	84.83 ± 2.85	97.67 ± 1.60	89.75 ± 1.74	97.63 ± 1.84
Original (flip)	87.50 ± 2.62	98.85 ± 0.44	93.46 ± 1.83	98.85 ± 0.77
(a) Rotation	92.69 ± 2.22	98.27 ± 0.38	93.08 ± 0.63	**99.04 ± 0.38**
(b) Blur	88.65 ± 1.31	98.65 ± 0.74	93.59 ± 1.94	98.85 ± 0.99
(c) Contrast	92.69 ± 0.44	**99.04 ± 0.38**	93.65 ± 0.74	98.65 ± 0.74
(d) Scaling	89.81 ± 0.74	**99.04 ± 0.97**	**95.00 ± 0.44**	98.65 ± 0.74
(e) Illumination	93.46 ± 2.84	98.46 ± 0.63	94.23 ± 0.99	**99.42 ± 0.38**
(f) Projective	93.08 ± 0.63	98.65 ± 0.74	93.65 ± 0.97	98.27 ± 1.31
(a) + (b)	92.50 ± 1.15	98.27 ± 0.38	93.27 ± 0.97	98.65 ± 1.15
(a) + (c)	95.00 ± 0.99	**99.04 ± 0.94**	94.81 ± 1.15	98.46 ± 0.89
(a) + (d)	92.69 ± 1.33	98.46 ± 0.63	93.65 ± 0.74	98.85 ± 1.33
(a) + (e)	**96.35 ± 0.74**	98.65 ± 1.31	94.42 ± 0.74	98.85 ± 1.33
(a) + (f)	92.69 ± 0.77	97.50 ± 0.97	93.65 ± 1.31	98.65 ± 0.74
(a) + (c) + (e)	**96.35 ± 0.97**	98.46 ± 0.63	94.23 ± 1.60	98.65 ± 0.74

4.2 AgrilPlant Dataset Evaluation

For the AgrilPlant dataset we also used the two CNN architectures trained from scratch or fine-tuned and evaluate them on both original and data-augmented datasets. The results are shown in Table 3. We observe that the fine-tuned GoogleNet with the combined effect of rotation and contrast yields the highest classification accuracy of 98.6%. The fine-tuned AlexNet profits most from the illumination DA technique. The performance improvements using DA on this dataset are much smaller than for the previous dataset. The reason is that there are 210 training images per class in this dataset, whereas there are only 14 training images per class in the Folio dataset. Still, for scratch AlexNet the combined DA techniques rotation+contrast and rotation+contrast+illumination result in a performance improvement of 2% compared to training from the original dataset. We also note that all CNN architectures with the blur DA technique obtain lower performances than using the original images. The reason is most probably that blurred images reduce the amount of salient features in the images from this dataset, which are still present in the test images.

4.3 Swedish Dataset Evaluation

The plant classification accuracies with different DA schemes on the Swedish dataset are reported in Table 4. The results show that the scratch CNN architectures profit from almost all DA methods. The biggest performance improvement is for scratch AlexNet where the use of the combined rotation+projective DA technique leads to a performance improvement of 3.1%. For this dataset, the fine-tuned CNN models do not profit from the DA techniques, and often the results using a DA technique are even a bit lower than using the original

Table 3. Recognition results using different DA schemes for the AgrilPlant dataset.

Augmentation methods	AlexNet		GoogleNet	
	Scratch	Fine-tuned	Scratch	Fine-tuned
Original [19]	89.53 ± 0.61	96.37 ± 0.83	93.33 ± 1.24	98.33 ± 0.51
(a) Rotation	90.10 ± 1.08	96.90 ± 0.69	92.53 ± 1.49	98.17 ± 0.68
(b) Blur	82.97 ± 2.26	94.43 ± 1.33	87.80 ± 1.27	97.73 ± 0.95
(c) Contrast	89.53 ± 1.26	96.27 ± 1.15	94.10 ± 0.95	98.17 ± 0.63
(d) Scaling	90.20 ± 0.95	96.93 ± 0.93	94.00 ± 1.20	98.13 ± 0.62
(e) Illumination	90.13 ± 1.06	**97.27 ± 0.38**	**95.03 ± 1.11**	98.21 ± 0.89
(f) Projective	90.87 ± 1.14	96.20 ± 0.92	93.21 ± 1.04	98.21 ± 0.76
(a) + (b)	87.70 ± 1.25	96.23 ± 0.71	90.40 ± 1.87	98.27 ± 0.62
(a) + (c)	**91.57 ± 0.96**	**97.10 ± 0.43**	**95.17 ± 1.38**	**98.60 ± 0.38**
(a) + (d)	90.40 ± 1.12	96.50 ± 0.31	92.93 ± 1.89	98.10 ± 0.82
(a) + (e)	**91.07 ± 0.49**	**97.03 ± 0.49**	94.07 ± 1.46	98.43 ± 0.60
(a) + (f)	90.50 ± 0.63	96.77 ± 0.95	92.77 ± 1.38	98.13 ± 0.92
(a) + (c) + (e)	**91.53 ± 0.78**	96.77 ± 0.71	94.73 ± 0.69	98.53 ± 0.59

dataset. The fine-tuned AlexNet obtained the best performance with the combined rotation+contrast+illumination DA technique and obtained an accuracy of 99.76%, while the fine-tuned GoogleNet worked best with the combined rotation+scaling DA method with an accuracy of 99.92%. Both these fine-tuned versions outperformed the previous study in [26] which combined shape, color, and Haralick texture features and reported an accuracy of 99.5%.

Table 4. Average accuracies and standard deviation using different DA techniques on the Swedish dataset.

Augmentation methods	AlexNet		GoogleNet	
	Scratch	Fine-tuned	Scratch	Fine-tuned
Original	94.69 ± 1.18	99.65 ± 0.07	96.08 ± 1.10	99.81 ± 0.15
(a) Rotation	96.21 ± 0.80	99.52 ± 0.29	97.04 ± 0.66	99.87 ± 0.13
(b) Blur	94.75 ± 0.97	99.36 ± 0.22	96.27 ± 1.40	99.57 ± 0.58
(c) Contrast	95.09 ± 0.67	99.55 ± 0.37	96.69 ± 0.91	99.79 ± 0.18
(d) Scaling	94.88 ± 0.78	99.60 ± 0.19	96.53 ± 1.25	99.84 ± 0.15
(e) Illumination	95.23 ± 0.53	99.49 ± 0.42	96.05 ± 0.91	99.76 ± 0.17
(f) Projective	96.88 ± 0.24	99.41 ± 0.07	96.64 ± 0.65	99.73 ± 0.13
(a) + (b)	96.40 ± 1.20	99.49 ± 0.17	97.12 ± 0.33	99.81 ± 0.15
(a) + (c)	97.07 ± 0.52	99.41 ± 0.15	**98.24 ± 0.52**	99.84 ± 0.11
(a) + (d)	96.40 ± 0.72	99.65 ± 0.22	97.68 ± 0.37	**99.92 ± 0.07**
(a) + (e)	97.25 ± 0.41	99.65 ± 0.28	98.16 ± 0.57	99.81 ± 0.22
(a) + (f)	**97.81 ± 0.77**	99.41 ± 0.07	97.55 ± 0.92	99.81 ± 0.07
(a) + (c) + (e)	97.60 ± 0.57	**99.76 ± 0.20**	97.68 ± 0.77	99.73 ± 0.16

4.4 Discussion

We have performed experiments on 3 datasets with 52 different techniques. If we look at the combined results, we can derive the following conclusions:

- The scratch version of AlexNet profits most from data augmentation. The reason is probably that it consists of most parameters to train and therefore larger datasets are very helpful.
- The fine-tuned CNN models hardly profit from data augmentation for the considered datasets. One reason is that the performances of the fine-tuned CNN methods are already very good, so there is not much room for improvement. Still, scaling helps the fine-tuned AlexNet with 0.3% average accuracy improvement and the illumination DA technique helps the fine-tuned GoogleNet a bit with an average accuracy improvement of 0.2%.
- For scratch AlexNet particular combined DA techniques lead to the biggest performance improvements. The average improvement for the three datasets using the combined DA technique rotation+contrast+illumination is 4.6%. This is followed by the combination of rotation and illumination with an average gain of 4.3%.
- For scratch AlexNet, the best single DA technique uses the projective transformation, which helps to improve the average accuracies with 3.0%.
- The scratch GoogleNet also profits most from combined DA techniques, where the combination of rotation and contrast helps to get 1.8% higher average accuracy.
- For scratch GoogleNet, the best single DA technique uses the scaling operator, which helps to improve the average accuracies with 0.9%.

5 Conclusion

We have investigated the usefulness of 6 different data-augmentation techniques and combinations of them using two well-known CNN architectures on three plant datasets. The results show that data-augmentation methods are important to obtain higher accuracies for CNN models trained from scratch. This shows that more training data helps a lot, which is also because some of the datasets do not contain many original training images. For the scratch AlexNet and GoogleNet architectures, especially the combined effects of rotation and illumination or rotation and contrast are very helpful. The blur operation does not help to obtain higher accuracies and sometimes even results in worse performances, despite the increase in the amount of training images. The fine-tuned AlexNet architecture profits a bit from the scaling DA technique, whereas the fine-tuned GoogleNet profits a bit from the illumination DA technique, but most other DA techniques are not helpful to obtain higher accuracies with the pre-trained CNN architectures. One reason why the fine-tuned CNN models do not really profit from data augmentation, is that they obtain very high performances on the considered datasets when trained on the original datasets. Therefore, there is very little room for improvement. The scratch CNN architectures in general

need much more training examples, and therefore profit a lot from the combined DA techniques which increase the number of different training images the most.

In future work, we want to examine the effects of data augmentation on more complex datasets for which the fine-tuned CNN architectures do not perform very well using only the original images. We also want to examine the data augmentation techniques describes in [18], where new images containing multiple version of an original image are constructed. Finally, instead of presetting the boundaries of the effects of the DA techniques, we want to focus on learning the right amounts in which images are changed with particular DA techniques using a novel adversarial learning framework.

References

1. Atabay, H.A.: A convolutional neural network with a new architecture applied on leaf classification. The IIOAB Journal 7, 226–231 (2016)
2. Bama, B.S., Valli, S.M., Raju, S., Kumar, V.A.: Content based leaf image retrieval (CBLIR) using shape, color and texture features. Indian Journal of Computer Science and Engineering 2(2), 202–211 (2011)
3. Dyrmann, M., Karstoft, H., Midtiby, H.S.: Plant species classification using deep convolutional neural network. Biosystems Engineering 151, 72–80 (2016)
4. Ghazi, M.M., Yanikoglu, B., Aptoula, E.: Plant identification using deep neural networks via optimization of transfer learning parameters. Neurocomputing (2017)
5. Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., Lew, M.S.: Deep learning for visual understanding: A review. Neurocomputing 187, 27–48 (2016)
6. Hsiao, J.K., Kang, L.W., Chang, C.L., Lin, C.Y.: Comparative study of leaf image recognition with a novel learning-based approach. In: Science and Information Conference (SAI), 2014. pp. 389–393. IEEE (2014)
7. Kadir, A., Nugroho, L.E., Susanto, A., Santosa, P.I.: Neural network application on foliage plant identification. arXiv preprint arXiv:1311.5829 (2013)
8. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. pp. 1097–1105 (2012)
9. Kumar, N., Belhumeur, P.N., Biswas, A., Jacobs, D.W., Kress, W.J., Lopez, I.C., Soares, J.V.: Leafsnap: A computer vision system for automatic plant species identification. In: Computer Vision–ECCV 2012, pp. 502–516. Springer (2012)
10. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature 521(7553), 436–444 (2015)
11. Lee, S.H., Chang, Y.L., Chan, C.S., Remagnino, P.: Plant identification system based on a convolutional neural network for the lifeclef 2016 plant classification task. In: Working notes of CLEF 2016 conference (2016)
12. Lin, M., Chen, Q., Yan, S.: Network in network. arXiv preprint arXiv:1312.4400 (2013)
13. McFee, B., Humphrey, E.J., Bello, J.P.: A software framework for musical data augmentation. In: ISMIR. pp. 248–254 (2015)
14. Mohanty, S.P., Hughes, D.P., Salathé, M.: Using deep learning for image-based plant disease detection. CoRR abs/1604.03169 (2016)
15. Mouine, S., Yahiaoui, I., Verroust-Blondet, A.: A shape-based approach for leaf classification using multiscale-triangular representation. In: Proceedings of the 3rd

ACM conference on International conference on multimedia retrieval. pp. 127–134. ACM (2013)
16. Munisami, T., Ramsurn, M., Kishnah, S., Pudaruth, S.: Plant leaf recognition using shape features and colour histogram with K-nearest neighbour classifiers. Computer Vision and the Internet (VisionNet'15), Second International Symposium on, Procedia Computer Science 58, 740 – 747 (2015)
17. Nilsback, M.E., Zisserman, A.: Automated flower classification over a large number of classes. In: Computer Vision, Graphics & Image Processing, 2008. ICVGIP'08. Sixth Indian Conference on. pp. 722–729. IEEE (2008)
18. Okafor, E., Smit, R., Schomaker, L., Wiering, M.: Operational data augmentation in classifying single aerial images of animals. In: Proceedings of the IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA) (2017)
19. Pawara, P., Okafor, E., Surinta, O., Schomaker, L., Wiering, M.: Comparing local descriptors and bags of visual words to deep convolutional neural networks for plant recognition. In: Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2017). pp. 479–486 (2017)
20. Salamon, J., Bello, J.P.: Deep convolutional neural networks and data augmentation for environmental sound classification. arXiv preprint arXiv:1608.04363 (2016)
21. Sato, I., Nishimura, H., Yokoi, K.: Apac: Augmented pattern classification with neural networks. arXiv preprint arXiv:1505.03229 (2015)
22. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
23. Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., Stefanovic, D.: Deep neural networks based recognition of plant diseases by leaf image classification. Computational Intelligence and Neuroscience 2016, 1–11 (2016)
24. Söderkvist, O.: Computer vision classification of leaves from swedish trees (2001)
25. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–9 (2015)
26. VijayaLakshmi, B., Mohan, V.: Kernel-based PSO and FRVM: An automatic plant leaf type detection using texture, shape, and color features. Computers and Electronics in Agriculture 125, 99–112 (2016)
27. Wang, X., Liang, J., Guo, F.: Feature extraction algorithm based on dual-scale decomposition and local binary descriptors for plant leaf recognition. Digital Signal Processing 34, 101–107 (2014)
28. Wang, Z., Lu, B., Chi, Z., Feng, D.: Leaf image classification with shape context and sift descriptors. In: Digital Image Computing Techniques and Applications (DICTA), 2011 International Conference on. pp. 650–654. IEEE (2011)
29. Zhang, S., Lei, Y., Zhang, C., Hu, Y.: Semi-supervised orthogonal discriminant projection for plant leaf classification. Pattern Analysis and Applications 19(4), 953–961 (2016)