Experimental evaluation of Data Augmentation heuristics for plant identification systems based on Deep Learning

Luciano Araújo Dourado Filho[1], Rodrigo Tripodi Calumby[1]

[1]University of Feira de Santana – Feira de Santana – BA – Brazil

`lucianoadfilho@ecomp.uefs.br, rtcalumby@uefs.br`

Abstract. *Data augmentation (DA) allows increasing datasets for training machine learning models that demands large amounts of data. In real-world applications in which data may not be abundant enough and data acquisition is not easy, DA enables increasing diversity and introducing model generalization. In this work we evaluate several DA techniques and combining approaches to extend image datasets used to train plant species recognition models. We experimentally validated Deep Convolutional Neural Networks (DCNN) with several datasets obtained from common augmentation techniques and combinations. The results allowed the identification of the Translate + Crop augmentation policy as the most effective within the scope of evaluation.*

1. Introduction

For a long time, machine learning algorithms for computer vision tasks were not effective or efficient enough for real-world applications, especially in the plant domain. Although innovative, these technologies mostly involved hand-craft designed solutions for single-organ detection (e.g., flower or leaves), recognition or segmentation from images. Therewith, ineffective models that were highly expensive to design and with low generalization power were a common reality. Hence, the use of those methods in critical systems (i.e., agriculture, pharmacology) was out of question. In turn, with the exploitation of high-end parallel architectures, those traditional solutions were considerably outperformed by *Deep Learning* (DL) models, specially the ones based on Deep Convolutional Neural Networks (DCNN) [Mehdipour Ghazi et al. 2017]. These advances allowed large-scale experiments and even the construction of models that surpassed human performance in some plant recognition tasks [Goëau et al. 2018].

The emergence of effective models and efficient models allowed the exploration of real-world experiments and the proposal of solutions for increasing and automatizing field production through DCNN-based systems in plant recognition tasks [Khan et al. 2018,

Brahimi et al. 2018, Zhao et al. 2020]. Despite that, the predictive models commonly include millions of parameters, demanding massive amounts of data for proper optimization, which still make the learning process unfeasible in some contexts. Besides that, additional challenges emerge from datasets with imbalanced class distributions or from the high cost for large-scale data acquisition. Such circumstances motivate researchers to use Data Augmentation (DA) for training DCNNs. DA allows to artificially extend a dataset through label-preserving transformations for the generation of novel samples from real ones. Indeed, DA has been reported as promising in overcoming dataset imbalance, reducing overfitting and introducing invariance in DL models [Shorten e Khoshgoftaar 2019].

In this context, recent works on automatic plant recognition through DCNN have commonly relied on DA [Sulc et al. 2018, Haupt et al. 2018], which shows a crucial role for achieving state-of-the-art classification. However, DA techniques have been used in an ad-hoc way, given the lack of a consensus regarding the suitability of the alternatives with respect to each context of application (i.e., organ type, photo type, background, herbarium sheet, etc). Besides, arbitrary costful heuristics can lead to performance bottlenecks and effectiveness decrease. Therefore, this work proposes a performance analysis of DA techniques commonly used for the optimization of plant species recognition models based on DL. Additionally, we assessed three heuristics for the combination of augmentation techniques, which allowed demonstrating the most effective techniques as well as the most suitable to perform plant species recognition.

2. Related Work

Data Augmentation enables to increase a dataset quantity in an artificial way [Buda et al. 2018]. The motivation behind that is to improve the DCNNs generalization power, through a invariance hardcoding procedure [Mehdipour Ghazi et al. 2017] that happens behind the process of training with images that presents variations of angle, position, light, brightness, simultaneously with the original corresponding versions. Consequently, the models become more invariant to these adversities and consequently, increase their capacity of performing well for unseen data [Shorten e Khoshgoftaar 2019, Mehdipour Ghazi et al. 2017].
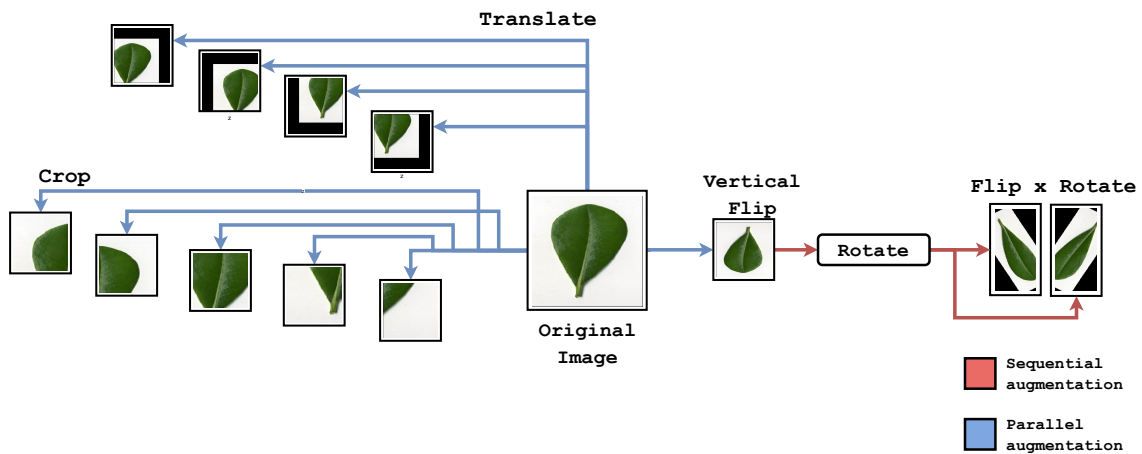

Figure 1. Illustration of geometric augmentation techniques.

Basically, augmentation techniques can be classified as geometric when they alter the geometric constitution of the images (Figure 1), or photometric when they only cause

changes over the color space [Taylor e Nitschke 2018]. Traditional geometric techniques includes: Crop, Flip (horizontally, vertically), Translate, Rotate, whereas photometric transformations mainly involves changes in brightness, light, color or saturation for example [Shorten e Khoshgoftaar 2019]. In terms of heuristics, these techniques (geometric or photometric) are usually applied individually or combined through sequential or parallel application heuristics to yield augmentation policies. In the *Sequential* modality, for example, samples are obtained from the chained application of multiple augmentation techniques over the same image. Figure 1 illustrates the sequential heuristic with a Flip x Rotate example. This process generates two rotated and vertically-flipped images, arising from the application of Flip followed by Rotate. In turn, the *Parallel* approach yields augmentation policies from the individual (parallel) application of two or more techniques and addition of the obtained images to a unique set. Besides that, some researchers have also proposed to use more complex heuristics, such as through Deep Learning based methods [Wang et al. 2017].

In the plant recognition context, the work in [Pawara et al. 2017] conducted an evaluation of augmentation techniques including Rotation, Blur, Contrast, Scaling, Illumination and Projection combined through a parallel approach. For that, three datasets were used: Folio (600 leaf images from 32 species), Swedish (1.000 images on a plain background of 15 Swedish tree species) and Agril Plant (3.000 images of 10 fruit species). Although promising results were achieved, it was not possible to draw generalizable techniques, possibly due to the low complexity of the datasets, with perfectly balanced classes, high background homogeneity (except for Agril Plant) and the small number of species/specimens.

In [Zhang et al. 2015] the authors proposed a CNN architecture for leaf classification and used Flavia dataset (1907 images of 32 species). The model was trained with traditional DA techniques (Translate, Scale, Rotate, Contrast and Sharpening) randomly selected according to a desired augmentation factor (5x, 10x or 20x). In a similar way, the authors observed a reasonable effectiveness improvement with DA, although, no expressive difference was observed for the techniques assessed and the low complexity of the dataset weakens the conclusion regarding the general effectiveness of the augmentation approaches. Similarly, in [Pandian et al. 2019] the authors assessed many traditional DA techniques, including Flip, Rotate, Crop, color transformation, PCA and noise injection, as well as DL-based techniques (WGAN, DCGAN, neural style transfer). An imbalanced plant dataset was used with around 54 thousand specimens from 38 classes including healthy and diseased leaves. The authors demonstrated that using all augmentation techniques (WGAN, DCGAN, neural style transfer and traditional techniques) at once lead to higher performance in relation to the use in isolation. Despite its complexity, higher costs and need for optimization, the DL-based augmentations applied in parallel allowed slightly superior results in contrast to the traditional techniques also combined in parallel.

As reported in [Mehdipour Ghazi et al. 2017], the combination by sequential and parallel application of vertical flip, rotation, and scaling augmentation techniques through a image patch extraction pipeline, allowed to achieve state-of-the-art performance in a multi-organ, large-scale plant classification task involving about 1,000 species. The authors observed a decrease of overfitting and the improvement of the benefits of fine-tuning. They also showed that an 80-fold augmentation outperformed a 10-fold by roughly 6% in

accuracy. This illustrates by how much the authors needed to increase the augmentation factor (no. of generated images) in order to obtain satisfactory results as well as how some augmentation heuristics can generate large datasets that increase training costs.

In general, large-scale studies that relied on DA demonstrated its important role towards developing real-world plant recognition systems. Nevertheless, most studies were not successful on determining general techniques or combining heuristics, mainly due to the low complexity of the data or the limited amount of combinations of DA techniques assessed. Hence, this work aims at assessing several DA techniques through different combination heuristics and a representative scenario of large-scale plant recognition in terms of class imbalance, visual image heterogeneity (multiple plant organs), number of species and specimens. Therefore, we believe to provide augmentation policies that are more suitable to the development of plant species recognition systems based on DCNN.

3. Experimental Setup

As discussed in [Pawara et al. 2017], the characteristics of the classification task and data used are decisive for robust DA studies involving DCNN. On the one hand, smaller datasets makes it difficult to conduct in-depth analysis of real-world challenges given the limited amount of samples and/or classes and its irregular distribution. On the other hand, datasets of larger scale/complexity makes comprehensive systematic analysis infeasible, specially considering the multiple possible combinations of DA techniques.

Considering the mentioned constraints, this work was conducted over the Plant-CLEF2013 (PC2013) dataset [Goëau et al. 2013]. It presents 26,077 images of 250 plant species from the French flora, from which 5092 (20%) images are used for testing. Introducing additional complexity, the dataset includes images acquired with different methods and of multiple plant organs, resulting in two image types and 6 sub-types as illustrated in Figure 2. The two main types are: *Sheet as Background* (a homogeneous surface of uniform white background) representing 42% of the total images (11,031 samples) and *Natural Background*, with 15,046 samples representing 58% of total images, composed of natural photographs captured outdoors from different perspectives of different organs.

Figure 2. Samples of images with multiple organs of different species divided by the categories "Sheet as Background" and "Natural Background".

For species recognition task we relied on transfer learning based on the ResNet [He et al. 2016] DCNN architecture[1]. In the preliminary experiments, ResNet50, ResNet101 and ResNet152 were fine-tuned with the individual application of each geometric augmentation technique: Translate, Crop, Flip and Rotate. The main goal of this phase was to assess the impact of the depth and the number of pre-adjusted (non-trainable) ImageNet weights over the results in order to select the most promising architecture for further analyzes. For that, we compared the performance of the baseline (training without augmentation) versus training also with the images generated by the DA technique. The preliminary experiments enabled selecting the better performing architecture to conduct the extended analysis. In this further analysis, we assessed 10 additional DA policies obtained through the sequential and parallel heuristics.

For an strict effectiveness assessment, the performance comparison relied on an independent test dataset. For test images, the same data augmentation policies were applied. Considering recognition is performed for 1+N images (1 original + N augmented versions), the class prediction is performed according to average Softmax scores. A summary of the DA policies assessed and the resulting amount images and augmentation factor are presented in Table 1. For these techniques, the following configurations were considered: Translate (4 different directions with offsets equivalent to 20% of the image width over the horizontal axis and 20% of the height over the vertical axis with black pixel padding); Rotate (30 degrees clockwise and counterclockwise); Crop (four corner patches and a central crop with 50% of the image size).

Table 1. Augmentation policies and resulting number of images. Bold values indicates the size in relation to the original dataset. *Augmentation factor in relation to the original dataset.

Dataset	Train+Validation	Test Set (*)	Dataset	Train+Validation	Test Set (*)
Original Dataset	20.985	5092	Flip + Translate	125.910	30.552 (**6x**)
Flip	41.970	10.184 (**2x**)	Flip + Crop	146.895	35.644 (**7x**)
Rotate	62.955	15.276 (**3x**)	Translate + Rotate	146.895	35.644 (**7x**)
Flip x Rotate	62.955	15.276 (**3x**)	Crop + Rotate	167.880	40.736 (**8x**)
Flip + Rotate	83.940	20.368 (**4x**)	Translate x Rotate	188.865	45.828 (**9x**)
Translate	104.925	25.460 (**5x**)	Translate + Crop	209.850	50.920 (**10x**)
Crop	125.910	30.552 (**6x**)	Translate x Crop	440.685	106.932 (**20x**)
Flip x Crop	125.910	30.552 (**6x**)			

The evaluation was conducted with a stratified random sampling protocol to keep class proportion in training and validation sets. More specifically, we considered 80% of the training data for model construction and the 20% for validation, followed by the testing with a an independent held-out test set. To ensure comparability, instead of performing individualized optimizations, all models were trained with the same configurations, for a fixed number of epochs (75) and amount of trainable weights. The Categorical Crossentropy loss function and the Accuracy measure were considered. On the test phase, the Micro-F1 measure was also computed to account for class imbalance. The weight update strategy used was the SGD with the Adam optimizer [Kingma e Ba 2017] and a Batch size of 64. First and Second Moment exponential decays and Epsilon were set to default (0.9, 0.99 and 10^{-7}, respectively), and the learning rate was set to 2×10^{-6}.

---

[1]ResNet is a recognized effective network, having achieved the first place at the ILSVRC 2015 classification task with 3.57% error on the ImageNet test set [Russakovsky et al. 2015].

4. Results And Discussions

At the preliminary step, we tested the models trained over the individual application of Crop, Translate, Flip and Rotate techniques. The results are presented in Table 2. The models trained over the augmented datasets yielded expressively superior results in comparison to the baseline. It demonstrates how promising these techniques are, with emphasis to the Crop and Translate techniques, allowing Micro-F1 over 0.46 with the ResNet152 model. Considering the network depths (50, 101 and 152 layers), increasing the number of pre-trained weights did not demonstrated to significantly affect the results. Despite the ResNet152 presented overall higher performance, it was not expressive enough in relation to the 50 and 101 variations. Therefore, the ResNet50 was chosen for the further analysis, since it presents significantly lower computational burden in comparison to ResNet101 and ResNet152 [He et al. 2016].

Table 2. Preliminary Results (Micro-F1 on test set).

Network	Baseline	Crop	Translate	Flip	Rotate
ResNet50	0.3177	0.4636	0.4550	0.4094	0.4049
ResNet101	0.3157	0.4693	0.4705	0.4108	0.4065
ResNet152	0.3189	0.4738	0.4652	0.4049	0.4102

For a deeper analysis with the ResNet50, we assessed 10 extra combinations of augmentation policies considering the parallel and sequential heuristics. The results presented in Table 3 demonstrates that DA led to equivalent or superior performance in relation to the baseline. In terms of heuristics, the parallel application policies demonstrated superiority in relation to the sequential and individual approaches. Regarding the sequential application, for example, the overall worst performing policy (Flip + Rotate, Micro-F1=0.4255) enabled similar results to the best performing Sequential policy (Flip x Crop, Micro-F1=0.4271). Considering that the chained (sequential) application of multiple geometric augmentation policies increases the chances of violating the label-preservation principle of data augmentation, we believe that this effect degraded the performance of the models trained with data generated with that heuristic.

Table 3. ResNet50 results: Combination heuristics, Augmentation Policies and performance (Micro-F1 on test set).

Baseline	
Augmentation	**Micro-F1**
None (Original data)	0.3177
Individual Application	
Crop	**0.4636**
Translate	0.4550
Flip	0.4094
Rotate	0.4049
Sequential Application	
Flip x Crop	**0.4271**
Translate x Rotate	0.4010
Flip x Rotate	0.3800
Translate x Crop	0.3189

Parallel Application	
Augmentation	**Micro-F1**
Translate + Crop	**0.4919**
Flip + Crop	0.4672
Crop + Rotate	0.4573
Translate + Rotate	0.4522
Flip + Translate	0.4518
Flip + Rotate	0.4255

Considering the number of resulting images (Table 1), our findings demonstrate that the DA policies that yielded better results not necessarily were the ones that presented higher augmentation factor. The *Translate x Crop* policy, for example, increased the original dataset to over 440.000 images (20x factor) and yet presented only 0.37% gain in relation to the Baseline. In contrast to that, the single application of Flip (2x factor) enabled improving classification performance by roughly 28%. In general, the combination of techniques that presented reasonable individual performance through parallel heuristics allowed superior results in comparison to the individual and sequential.

5. Conclusions

Data augmentation was investigated in this work for leveraging DCNN models for plant species recognition from images. It included a broad assessment of data augmentation policies and combination heuristics for the mostly frequent techniques in the literature. Despite the difficulties behind the process of classification from imbalanced, heterogeneous multi-organ plant datasets, the most effective augmentation policies evaluated allowed expressive results with significant performance gains.

Our results can guide future development of DCNN-based plant recognition systems from imbalanced datasets. We believe that systems like these can play an important role in the development of novel production models, capable of improving precision and automatizing several tasks in fields such as agriculture, pharmacology, ecology, environmental engineering, etc. Finally, we believe novel experiments can be conducted to more deeply analyze the DA potential on overcoming class imbalance.

6. Acknowledgement

References

Brahimi, M., Arsenovic, M., Laraba, S., Sladojevic, S., Boukhalfa, K., e Moussaoui, A. (2018). *Deep Learning for Plant Diseases: Detection and Saliency Map Visualisation*, pages 93–117. Springer International Publishing, Cham.

Buda, M., Maki, A., e Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 106:249 – 259.

Goëau, H., Bonnet, P., e Joly, A. (2018). Overview of expertlifeclef 2018: how far automated identification systems are from the best experts?

Goëau, H., Joly, A., Bonnet, P., Bakic, V., Barthélémy, D., Boujemaa, N., e Molino, J.-F. (2013). The imageCLEF Plant Identification Task 2013. *Proceedings of the 2nd ACM International Workshop on Multimedia Analysis for Ecological Data*, (i):23–28.

Haupt, J., Kahl, S., Kowerko, D., e Eibl, M. (2018). Large-scale plant classification using deep convolutional neural networks. In *CLEF (Working Notes)*.

He, K., Zhang, X., Ren, S., e Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Khan, M. A., Akram, T., Sharif, M., Awais, M., Javed, K., Ali, H., e Saba, T. (2018). Ccdf: Automatic system for segmentation and recognition of fruit crops diseases based on correlation coefficient and deep cnn features. *Computers and electronics in agriculture*, 155:220–236.

Kingma, D. P. e Ba, J. (2017). Adam: A method for stochastic optimization.

Mehdipour Ghazi, M., Yanikoglu, B., e Aptoula, E. (2017). Plant identification using deep neural networks via optimization of transfer learning parameters. *Neurocomputing*.

Pandian, J. A., Geetharamani, G., e Annette, B. (2019). Data augmentation on plant leaf disease image dataset using image manipulation and deep learning techniques. In *2019 IEEE 9th International Conference on Advanced Computing (IACC)*, pages 199–204. IEEE.

Pawara, P., Okafor, E., Schomaker, L., e Wiering, M. (2017). Data augmentation for plant classification. In *International Conference on Advanced Concepts for Intelligent Vision Systems*, pages 615–626. Springer.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., e Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252.

Shorten, C. e Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):60.

Sulc, M., Picek, L., e Matas, J. (2018). Plant recognition by inception networks with test-time class prior estimation. In *CLEF (Working Notes)*.

Taylor, L. e Nitschke, G. (2018). Improving deep learning with generic data augmentation. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1542–1547. IEEE.

Wang, J., Perez, L., et al. (2017). The effectiveness of data augmentation in image classification using deep learning. *Convolutional Neural Networks Vis. Recognit*, 11.

Zhang, C., Zhou, P., Li, C., e Liu, L. (2015). A convolutional neural network for leaves recognition using data augmentation. In *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, pages 2143–2150. IEEE.

Zhao, B., Li, J., Baenziger, P. S., Belamkar, V., Ge, Y., Zhang, J., e Shi, Y. (2020). Automatic wheat lodging detection and mapping in aerial imagery to support high-throughput phenotyping and in-season crop management. *Agronomy*, 10(11).