LEAF CLASSIFICATION USING MARGINALIZED SHAPE CONTEXT AND SHAPE+TEXTURE DUAL-PATH DEEP CONVOLUTIONAL NEURAL NETWORK

Meet P. Shah[1], *Sougata Singha*[2], *Suyash P. Awate*[3*]

[1]Electrical Engineering Department, Indian Institute of Technology (IIT) Bombay. [2]Yodlee.
[3]Computer Science and Engineering Department, Indian Institute of Technology (IIT) Bombay.

ABSTRACT

Identifying plant species based on photographs of their leaves is an important problem in computer vision and biology. Previous approaches for leaf image classification typically rely on hand-crafted shape features or texture features. In contrast, we propose a *dual-path* deep convolutional neural network (CNN) to (i) learn *joint* feature representations for leaf images, exploiting their *shape* and *texture* characteristics, and (ii) optimize these features for the classification task. We compare our CNN approach against (i) vanilla CNN classifiers and (ii) popular hand-crafted shape features, including a novel shape-context based feature that is extremely computationally efficient, which we call the *marginalized shape context*. Our results on three large public datasets demonstrate that our dual-path CNN leads to higher accuracy and consistency than the state of the art.

Index Terms— Leaf recognition, shape, texture, marginalized shape context, dual-path deep convolutional neural net.

1. INTRODUCTION

Plant species can be identified through several geometrical and appearance characteristics exhibited in its roots, stem, leaves, flowers, and fruits. However, for many of these characteristics, the inter-species differences are often subtle and non-trivial to model mathematically and computationally. Nevertheless, plant leaves contain rich information for species identification because (i) leaves stay attached to the plants for longer durations, compared to fruits and flowers, and (ii) leaves have fairly distinctive characteristic features like shape and texture. This paper deals with recognizing plant species using an image of its leaf [1, 2, 3, 4, 5, 6, 7, 8].

We consider *shape* to be the geometrical information remaining in the boundary of an object after the information related to pose, location, and size is removed [9]. While leaf size changes with the age of the leaf, the shape of a leaf remains relatively unchanged over time and geographical location [6]. In addition to leaf shape, we exploit information in the leaf texture, subsuming venation and color information.

*Thanks to IIT Bombay Seed Grant 14IRCCSG010 and Nvidia GPU Grant. All work done at IIT Bombay.

Typical methods for leaf classification rely primarily on (i) leaf shape and (ii) hand-crafted features. For instance, the popular plant species identification system Leafsnap [4] uses curvature-based shape features at multiple scales, without using any color / texture features. Other methods capture shape characteristics via, e.g., moment invariants or the multiscale distance matrix [3]. Other approaches [2, 10] rely solely on texture information, ignoring shape information. Indeed, the information contained in leaf shape and texture is quite complementary. Recent approaches [11, 12] rely on a convolutional neural network (CNN) to classify leaf images or manually-selected patches from the leaf images. The manual patch selection ignores shape information and increases the sensitivity of the results to the manual input. Also, we show that by learning complementary features for shape and texture in separate CNN-architecture paths, but optimizing them jointly for classification, CNNs can learn features more effectively and improve recognition accuracy.

In this paper, we propose a *dual-path* deep convolutional neural network (CNN) to (i) learn *joint* feature representations for leaf images, exploiting their *shape* and *texture* characteristics, and (ii) optimize these features for the classification task. We compare our CNN approach against (i) vanilla CNN classifiers, like [11, 12], and (ii) several popular hand-crafted shape features. The results on 3 large public datasets demonstrate that our dual-path CNN leads to higher accuracy and consistency than the state of the art. In the context of comparing our dual-path CNN with hand-crafted shape features, we also propose a novel shape-context based feature that is extremely computationally efficient, which we call the *marginalized shape context*. Unlike popular descriptors like the shape context [13], the marginalized shape context does *not* require point correspondences or alignment across leaf shapes, making it fast and robust. Its low dimensionality and specific design add to its robustness. The marginalized shape context outperforms other popular hand-crafted shape features for leaf image classification.

2. RELATED WORK

Works on plant species identification can be categorized based on the features and the classifiers employed. For instance, (i) several methods rely solely on shape information [1, 3, 4], (ii) a few methods rely only on color and texture [2] (ignoring shape), and (iii) some methods rely on leaf shape and the vein geometry within the leaf [10, 7] (ignoring color and texture). [14] focus only on vein structure extraction (without classification). Recognition rates for [10] are quite low (around 0.6) for certain classes of leaves. The problem of vein detection [10, 7] relies on edge detection that is known to be ill posed in scenarios with low contrast-to-noise ratio, thereby reducing the robustness of the results to the choice of the free parameter values and heuristics, and with poor performance resulting from suboptimal choices.

Most methods using shape descriptors use a k-nearest neighbour or a support vector machine (SVM) as the underlying classifier. Recent classification methods [11, 12, 15] employ neural networks as the underlying classifier. The method in [15] employs a probabilistic neural network for classification, using 12 leaf features orthogonalized into 5 principal variables, yielding 90% classification accuracy on the Flavia dataset [16]. In contrast, our dual-path CNN exploits rich information in the entire leaf image and yields near-perfect recognition on the Flavia dataset. Lee et al. [11] employ a CNN [17] with inputs being leaf image patches capturing texture, but ignore global information like the leaf shape. [18, 19] use hand-crafted features of texture and shape. Zhao et al. [12] propose a CNN with two hidden layers, starting training from a simple structure of a single convolution kernel and gradually adding more convolution neurons to the architecture. Simultaneously, the growing connection weights are modified to minimize the squared-error. Unlike [12], our dual-path CNN learns complementary features for shape and texture in separate CNN-architecture paths, but optimize them jointly for classification. The results show that the CNNs in [11, 12] fail to generalize on larger leaf image datasets exhibiting more variation in shape and appearance.

3. METHODS

This section describes our dual-path CNN architecture (Figure 1), including schemes for data preprocessing, data augmentation, and hyperparameter optimization. It also describes a new hand-crafted shape feature, i.e., the marginalized shape context, which is robust and computationally efficient.

3.1. Dual-Path Convolutional Neural Network (CNN)

Dual-Path CNN Architecture. Our CNN architecture comprises two pathways, one for learning shape-dependent features and another for learning texture features. These paths later join to combine complementary information on shape and texture for optimal leaf image classification, into a multilayer perceptron. For a leaf image input to the CNN, we generate two different images, i.e., the **leaf-image** and the **texture-patch**, which are input to the two pathways of CNN

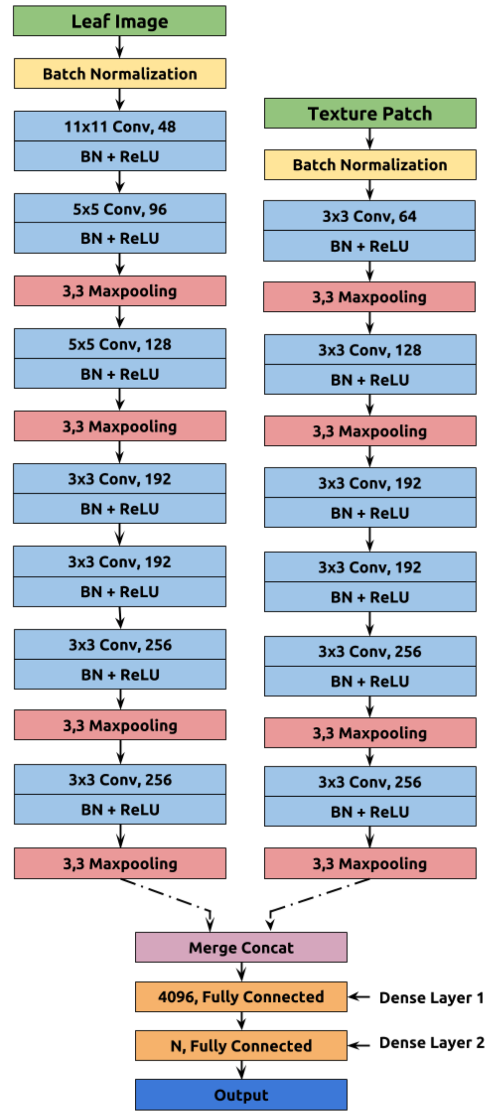

Fig. 1. Our **Dual-Path CNN**. "BN + ReLU" $\equiv$ batch normalization followed by ReLU activation; $N \equiv$ number of classes.

model. The original leaf image is first aligned to a common coordinate frame by registration, under a similarity transform, to a template leaf image. Then, the **leaf-image** input (Figure 2) is simply the entire leaf image resized (scaled down) to 144×192 pixels, which captures primarily shape (and some color) information. The **texture-patch** input (Figure 2) is obtained by 2x enlargement of the original leaf image, sharpening, and subsequent cropping of the central region of the leaf to get a patch of size 144×192 pixels; this primarily captures details relating to leaf texture and venation.

The feature extraction component of the CNN comprises convolutional and max-pooling layers. Each convolutional layer is followed by a batch normalization (BN) [20] layer. The activation used in all layers, except the final layer, is the rectified linear unit (ReLU) [21]. We concatenate the features learned from both pathways to form the joint shape-texture representation of the leaf. This joint representation is input to a multilayer perceptron for classification.

Dual-Path CNN Training. Data augmentation schemes improve image classification tasks that use neural networks by preventing overfitting and simulating more examples for learning [17]. Thus, we use randomized affine-transformed images to augment the data. We train by minimizing the batchwise categorical cross-entropy of the predicted and the true labels, using stochastic gradient descent. We use early stopping [22] to stop training when the validation loss begins to increase continuously for a given number of epochs. To optimize the hyperparameters (batch size, learning rate, learning rate decay, momentum in the gradient descent), we employ the covariance matrix adaptation evolution strategy [23].

Joint Feature Analysis. Deconvolution networks [24] help visualize and understand features learned by CNNs and can, thereby, help improve CNN architectures. Lee et al. [11] use a similar scheme to project the feature maps back to the input pixel space, by alternating deconvolution and unpooling layers, and visualize the features learned by the CNN for an input texture image. Since the input to their CNN is only patches, the features learned by the CNN for two leaves belonging to different species with similar texture but different shapes have similar activation maps. Indeed, their failure analysis [11] reveals that most leaf pairs misclassified by their CNN model belonged to leaf pairs with similar texture but different shapes. Our dual-path CNN exploits both shape and texture characteristics to perform classification and thus accurately classifies such leaf pairs. We analyze the joint feature representation learned by our proposed model, to show the capability of our method to recognize such leaf pairs (Figure 2, Table 1) and to establish shape as a crucial discriminatory factor, complementary to texture, in leaf classification. In Figure 2, while leaves A and B have similar shapes but different texture (and venation), leaves C and D have similar texture but different shapes. Our dual-path CNN is able to discriminate between these leaves (A versus B; C versus D) far better than texture-based CNN [11] and uni-path CNN [12].

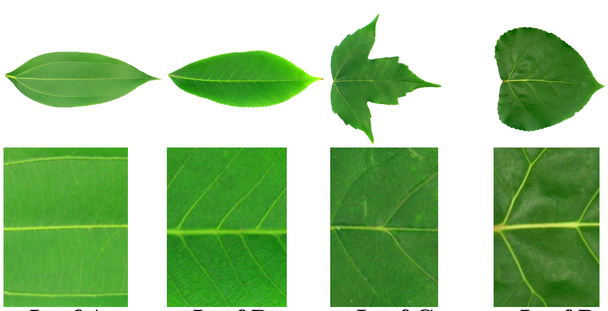

Fig. 2. Example Leaves Differing in Shape and Texture. *Top Row.* Leaf images input to the CNN path (left path in Figure 1) primarily capturing shape. *Bottom Row.* Texture patches input to the CNN path capturing texture.

Dissimilarity	Dual-Path CNN		Texture-Patch CNN [11]		Uni-Path CNN [12]	
	Dense-1	Dense-2	Dense-1	Dense-2	Dense-1	Dense-2
RRMSD: A-B	**0.73**	**0.30**	0.39	0.24	0.13	0.24
RRMSD: C-D	**0.85**	**0.37**	0.28	0.12	0.46	0.34
1-NCC: A-B	**0.13**	**0.10**	0.06	0.05	0.03	0.03
1-NCC: C-D	**0.09**	**0.11**	0.05	0.02	0.08	0.09

Table 1. Utility of Dual Shape+Texture Features. Dissimilarity in activations of dense layers 1 and 2 (Figure 1) to compare the discriminability of CNN architectures to leaves: (i) A versus B and (ii) C versus D. RRMSD $\equiv$ relative root mean squared difference. NCC $\equiv$ normalized cross correlation.

3.2. Hand-Crafted Feature: Marginalized Shape Context

We compare our dual-path CNN to hand-crafted shape features that have been very successful [4] for leaf image classification. In this context, we propose a new shape feature relying on the popular shape context [13].

We first get a set of points on the leaf boundary by (i) segmenting the image by converting it to grayscale followed by Gaussian convolution and thresholding, (ii) removing the leaf stem by morphological opening, and (iii) subsampling the boundary points at equal distances (measured via pixel counts). We design our shape descriptor to be: (i) invariant to the leaf's location by centering the coordinate system so that its origin coincides with the centroid of the pointset and (ii) invariant to scale by rescaling the coordinates of each point in the centered pointset by a constant factor so that the sum of squared distances of each point from the origin equals 1. To achieve rotational invariance, at each point in the pointset, we use tangent vectors to the leaf boundary (instead of a cardinal axis, along an image dimension) to define the angular coordinates underlying the shape context [13]. For R radial r-bins and T angular θ-bins, the shape-context histogram [13] of each point has dimension RT.

We take both the radial and angular bin limits to be linearly spaced, unlike log-linear spaced (i.e., linear spacing in log scale) radial bin limits in [13]. In the context of leaf classification, the main motivation for log-linear spaced radial bin limits is to alleviate the corruption of the descriptor from the stem present in the image [3]. However, since we eliminate the leaf stem during preprocessing, we replace the log-linear spacing with linear spacing to capture finer leaf details.

The conventional approach [13] concatenates the per-point local shape context vectors to form a shape descriptor for the entire object. If the number of points in the shape is N, the shape descriptor has dimension NRT. To evaluate the dissimilarity between two shape pointsets, the strategy in [13] requires point correspondences between them, which is very expensive having complexity $\mathcal{O}(N^3)$ [25]. In contrast, our approach eliminates the need for correspondence finding (making it orders of magnitude faster) while greatly improving performance over hand-crafted features.

In the shape-context histogram of a point p, the value at some r–θ bin represents the probability of a neighbouring point q being in that bin, for that shape. We propose to *marginalize* these probabilities across all points in the shape to get a new shape descriptor for the object boundary. This marginalized feature is orders-of-magnitude *faster* to compute because it does *not* require any correspondences across pointsets. Suppose we have R radial bins and T angular bins, our marginalized shape context descriptor has dimension RT, which is independent of boundary pointset cardinality N and, thereby, much smaller than the dimension NRT for the concatenated shape context descriptor in the traditional approach [13]. Furthermore, the marginalization, which is essentially an integration process, makes our descriptor more *robust to errors* in segmentation, point placement, and binning. Figure 3 shows our descriptors; the leaves on the left and the middle of the figure (belonging to the same plant species) have similar shape descriptors, unlike that of the leaf on the right (belonging to another plant species). This shows that our descriptor has the potential to model leaf shapes well.

4. RESULTS AND DISCUSSION

We evaluate 6 leaf classification methods: (i) our dual-path CNN, (ii) uni-path CNN [12], (iii) texture-patch CNN [11], (iv) our marginalized shape context with a SVM classifier (histogram intersection kernel [26]), (v) curvature histogram used by leafsnap [4] with a SVM classifier (histogram intersection kernel), and (vi) Multiscale distance matrix (MDM) [3] with a SVM classifier (Gaussian kernel). For multi-class classification using SVM, we use the one-versus-all technique [27] and tune parameters using 3-fold cross validation. We evaluate leaf classification methods on 3 large publicly available data sets: (i) Flavia Dataset (1907 images; 32 classes) [16], (ii) Leafsnap Dataset (7710 images; 150 classes) [28], and (iii) ImageClef dataset (6630 images; 126 classes) [29]. We split each dataset into two, taking a random subset of 70% leaves from each leaf class for training and 30% for testing. We repeat the random selection 20 times (bootstrap samnpling) and show the mean top-1 and top-3 classification accuracies (and the standard deviation) to depict the variability of the classification performance with respect to variability in the choice of the training and test sets. To evaluate the quality of classification, for each leaf in the test set, we compute the top-K classes ($K = 1, 2, 3$) and check if top K matches contains the true class for the leaf.

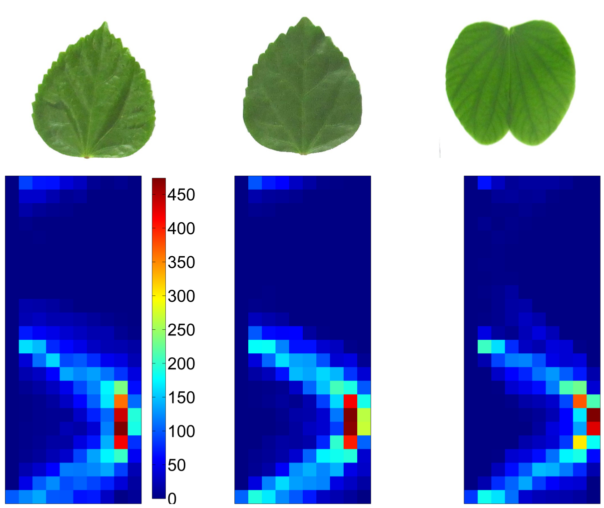

Fig. 3. **Marginalized Shape Context** for 3 leaves: the leaves on the left and the middle belong to the same species.

Method	Flavia		Leafsnap		ImageClef	
	Top-1	Top-3	Top-1	Top-3	Top-1	Top-3
Dual-Path CNN	**99.28**	**99.97**	**95.61**	**99.67**	**96.42**	**99.35**
	(0.18)	(0.10)	(0.24)	(0.18)	(0.30)	(0.24)
Uni-Path CNN [12]	96.22	99.46	88.83	97.37	84.21	94.40
	(0.26)	(0.21)	(0.58)	(0.21)	(0.34)	(0.22)
Texture-Patch CNN [11]	88.19	97.80	78.72	92.43	70.35	87.83
	(1.68)	(0.37)	(1.37)	(0.42)	(1.71)	(0.35)
Marginalized SC + SVM	*93.22*	*98.76*	*85.37*	*91.94*	*81.63*	*88.72*
	(1.05)	(0.33)	(0.23)	(0.22)	(0.28)	(0.35)
Curvature SC [4] + SVM	82.84	94.62	71.13	78.43	68.59	81.17
	(1.63)	(1.21)	(1.07)	(1.16)	(1.74)	(1.55)
MDM [3] + SVM	82.55	95.12	75.76	85.96	53.22	61.11
	(1.61)	(1.10)	(0.50)	(0.22)	(0.31)	(0.35)

Table 2. **Leaf Image Classification Performance.** SC $\equiv$ Shape Context. Top-K accuracy, with mean and standard deviation (in parenthesis) depicting variability (via bootstrap sampling) with respect to the choice of training and test sets.

Our proposed *dual-path CNN* outperforms all methods, giving near-perfect top-1-match results on the Flavia dataset and near-perfect top-3-match results on all datasets. The results also demonstrate that the dual-path architecture, enabling each path to specialize in one of the complementary shape and texture features, does better than the uni-path approach [12]. Moreover, the texture-patch based CNN [11] that ignores shape information performs the worst among CNN approaches, and even worse than our hand-crafted marginalized shape context, emphasizing the utility of shape features in leaf classification. The results show the benefits of automatic feature learning using CNNs, e.g., the dual-path and unipath-CNN, over all hand-crafted features. Among the hand-crafted features, all using the nonlinear SVM classifier, our *marginalized shape context* performs the best.

Conclusion. This paper presents a *dual-path CNN* for leaf image classification, which *jointly* learns complementary *shape* and *texture* characteristics in each path and optimizes them for leaf classification. We demonstrate the utility of shape and texture both for leaf classification. Our dual-path CNN outperforms the state of the art including CNN-based methods and hand-crafted features, on 3 large public datasets. We also propose a novel shape feature, i.e., the *marginalized shape context*, that is computationally efficient and outperforms popular hand-crafted features.

5. REFERENCES

[1] P Belhumeur, D Chen, S Feiner, D Jacobs, W Kress, H Ling, I Lopez, R Ramamoorthi, S Sheorey, and S White, "Searching the world's herbaria: A system for visual identification of plant species," in *Proc. Euro. Conf. Comp. Vis.*, 2008, pp. 116–129.

[2] J Cope, P Remagnino, S Barman, and P Wilkin, "Plant texture classification using Gabor co-occurrences," in *Proc. Adv. Vis. Comput.*, 2010, pp. 669–77.

[3] R Hu, W Jia, H Ling, and D Huang, "Multiscale distance matrix for fast plant leaf recognition," *IEEE Trans. Imag. Proc.*, vol. 21, no. 11, pp. 4667–4672, 2012.

[4] N Kumar, P Belhumeur, A Biswas, D Jacobs, W Kress, I Lopez, and J Soares, "Leafsnap: A computer vision system for automatic plant species identification," in *Proc. Euro. Conf. Comp. Vis.*, 2012, pp. 502–516.

[5] S Mouine, I Yahiaoui, and A Verroust-Blondet, "Advanced shape context for plant species identification using leaf image retrieval," in *Proc. Int. Conf. Multimedia Retrieval*, 2012, p. 49.

[6] S Mouine, I Yahiaoui, and A Verroust-Blondet, "A shape-based approach for leaf classification using multiscale triangular representation," in *Proc. Int. Conf. Multimedia Retrieval*, 2013, pp. 127–34.

[7] Y Nam, E Hwang, and D Kim, "A similarity-based leaf image retrieval scheme: Joining shape and venation features," *Comp. Vis. Imag. Understanding*, vol. 110, no. 2, pp. 245–59, 2008.

[8] Z Wang, Z Chi, and D Feng, "Shape based leaf image retrieval," *IEE Proc. Vis. Imag. Signal Proc.*, vol. 150, no. 1, pp. 34–43, 2003.

[9] I Dryden and K Mardia, *Statistical Shape Analysis*, John Wiley and Sons, 1998.

[10] J Park, E Hwang, and Y Nam, "Utilizing venation features for efficient leaf image retrieval," *J. Systems Software*, vol. 81, no. 1, pp. 71–82, 2008.

[11] S Lee, C Chan, P Wilkin, and P Remagnino, "Deepplant: Plant identification with convolutional neural networks," in *IEEE Int. Conf. Imag. Proc.*, 2015, pp. 452–6.

[12] Z Zhao, B Xie, Y Cheung, and X Wu, "Plant leaf identification via a growing convolution neural network with progressive sample learning," in *Proc. Asian Conf. Comp. Vis.*, 2014, pp. 348–61.

[13] S Belongie, J Malik, and J Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, 2002.

[14] J Clarke, S Barman, P Remagnino, K Bailey, D Kirkup, S Mayo, and P Wilkin, "Venation pattern analysis of leaf images," in *Int. Symp. Vis. Comput.*, 2006, pp. 427–36.

[15] S Wu, F Bao, E Xu, Y Wang, Y Chang, and Q Xiang, "A leaf recognition algorithm for plant classification using probabilistic neural network," in *Proc. Int. Symp. Sig. Proc. Info. Tech.*, 2007, pp. 11–16.

[16] "Flavia dataset," `flavia.sourceforge.net`, Accessed: 2017-01-01.

[17] A Krizhevsky, I Sutskever, and G Hinton, "Imagenet classification with deep convolutional neural networks," in *Adv. Neural Info. Proc. Systems*, 2012, pp. 1–8.

[18] J Chaki, R Parekh, and S Bhattacharya, "Plant leaf recognition using texture and shape features with neural classifiers," *Pattern Recog. Let.*, vol. 58, 61-68.

[19] T Kumar, V Reddy, and P Bora, "Leaf identification using shape and texture features," in *Int. Conf. Comp. Vis. Imag. Proc.*, 2016, pp. 531–541.

[20] S Ioffe and C Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–56.

[21] V Nair and G Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 807–814.

[22] R Caruana, S Lawrence, and C Giles, "Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping," in *Proc. Neural Info. Proc. Systems*, 2000, pp. 402–408.

[23] I Loshchilov and F Hutter, "CMA-ES for hyperparameter optimization of deep neural networks," *Int. Conf. Learning Represent.*, 2016.

[24] M Zeiler and R Fergus, "Visualizing and understanding convolutional networks," in *Proc. Euro. Conf. Comp. Vis.*, 2014, pp. 818–33.

[25] H Kuhn, "The Hungarian method for the assignment problem," *Naval Res. Logist.*, vol. 2, pp. 83–97, 1955.

[26] A Barla, F Odone, and A Verri, "Histogram intersection kernel for image classification," in *Proc. Int. Conf. Imag. Proc.*, 2003, vol. 3, pp. 510–3.

[27] B Scholkopf and A Smola, *Learning with Kernels*, MIT Press, 2002.

[28] "Leafsnap dataset," `leafsnap.com/dataset`, Accessed: 2017-01-01.

[29] "ImageClef 2012 dataset," `www.imageclef.org/2012/plant`, Accessed: 2017-01-01.