Evaluation of Features for Leaf Classification in Challenging Conditions

David Hall Chris McCool Feras Dayoub Niko Sünderhauf Ben Upcroft
ARC Centre of Excellence for Robotic Vision
Queensland University of Technology, 2 George St., Brisbane, Australia
d20.hall@qut.edu.au

Abstract

Fine-grained leaf classification has concentrated on the use of traditional shape and statistical features to classify ideal images. In this paper we evaluate the effectiveness of traditional hand-crafted features and propose the use of deep convolutional neural network (ConvNet) features. We introduce a range of condition variations to explore the robustness of these features, including: translation, scaling, rotation, shading and occlusion. Evaluations on the Flavia dataset demonstrate that in ideal imaging conditions, combining traditional and ConvNet features yields state-of-the-art performance with an average accuracy of $97.3\% \pm 0.6\%$ *compared to traditional features which obtain an average accuracy of* $91.2\% \pm 1.6\%$*. Further experiments show that this combined classification approach consistently outperforms the best set of traditional features by an average of* 5.7% *for all of the evaluated condition variations.*

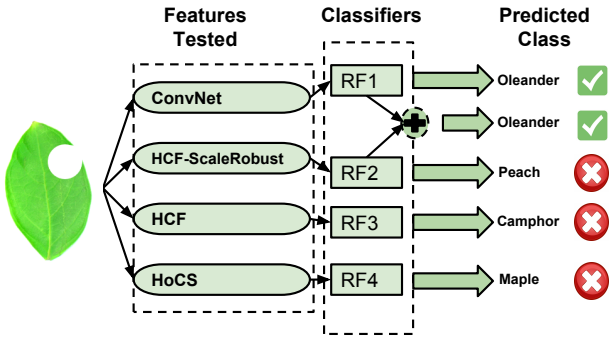

Figure 1: An overview of the experiments conducted in this paper. We introduce a variety of condition variations to the test dataset, like the occlusion shown on the leaf above. Several features are extracted from these altered images and passed through the random forest (RF) classifiers that were trained on perfect images. We compare the variability of the different tested systems and one of their combinations.

1. Introduction

Fine-grained image classification has received considerable attention in the past five years. The fine-grained image classification problem is often described as being able to perform species level classification, that is to tell the difference between an Elm tree and a Pine tree.

An area of particular interest for fine-grained image classification is leaf classification. There have been recent international competitions which have examined this topic, for instance the recent ImageCLEF challenge for plant classification (including leaf images) [11]. As well as this, leaf classification has several exciting practical applications such as in agriculture. In the agricultural setting, being able to distinguish between weeds and crop is of considerable interest as it allows for selective spraying [25] and allows for mapping of weed and crop distributions. Such an approach requires autonomous operation and means that the images will be captured in challenging environments. These challenging environments can have different effects on the appearance of a leaf such as the shape, colour and texture as well as changes due to viewpoint and occlusion.

Much of the research for leaf classification has concentrated on the use of traditional hand-crafted features for images taken in ideal conditions. Several databases have been proposed [26, 15] all of which consist of well aligned images taken in controlled conditions with few defects and often without a way to reproduce the experiments. The most often used features are hand-crafted features (in the following abbreviated as HCFs) that capture shape information [15, 17] or both shape and statistical information [8]. To date, limited work has been conducted on the use of learnt features such as those trained using deep convolutional neural networks (ConvNets) [16, 14, 10].

In this work, we examine the robustness of commonly used features for fine-grained leaf classification under challenging conditions. An extensive set of evaluations is performed by introducing a variety of condition variations such as: translation, scaling, rotation, shading and occlusion. We apply this evaluation to traditional HCFs as well the re-

cently proposed ConvNet features [10]. An overview of our testing procedure is shown in Figure 1. Empirical evaluation shows that the ConvNet features outperform the traditional HCFs and that their combination yields state-of-the-art performance with an average accuracy of $97.3\% \pm 0.6\%$ compared to $91.2\% \pm 1.6\%$ for the HCFs.

Two major contributions are made in this paper for leaf classification:

1. We examine the robustness of a range of features to a variety of condition variations including: translation, scaling, rotation, occlusion and shadowing.

2. We demonstrate that combining ConvNet features and HCFs leads to a robust state-of-the-art leaf classification system which has an average absolute performance improvement of 5.7% compared to the best performing HCF-based system.

The rest of the paper is organised as follows. In Section 2, we discuss related work in the field. Section 3 gives an overview for the approach used in this work. In Section 4 we present the experimental set-up and results for the various experiments conducted. Finally we draw conclusions and discuss future work in Section 5.

2. Related Work

Automatic leaf classification from imagery has been an active field of research since 1999. A range of hand-crafted features have been proposed to solve this problem ranging from shape-based features [26, 17, 2, 12, 18, 4, 19, 15] through to the use of statistical texture features [8] and even their spectral reflectance properties [20, 21].

Much of the work conducted on leaf classification makes use of shape features. Hand-crafted shape features such as the size, roundness, and convexity of the leaf have been used by several authors [26, 17, 2, 12, 18, 4]. Recently, more advanced shape features have also been proposed such as the triangular representation of Mouine et al. [19] and the histogram-based features of the well known LeafSnap[1] application of Kumar et al. [15]. Kumar et al. proposed a shape feature that measures the intersection between the segmented leaf and discs of varying sizes that were applied across the image. The different responses in terms of area of intersection and scale of the disc were then used to form a histogram of curvature scales (HoCS). This approach was designed to be robust to translations and rotations.

Other features for classification have come from the spectral properties of the leaves or from combining shape and statistical information. In [20, 21], the authors demonstrated that plant species could be classified by using hyperspectral images of their leaves. However, this preliminary work has concentrated on differentiating a particular crop and weeds and so relatively few classes had to be classified; in [20] six species were classified and in [21] eight species were classified.

[1] http://leafsnap.com/

Figure 2: Examples of different leaf species from the Flavia Dataset [26].

Work in 2014 by Haug et al. [8] combined shape and statistical features to distinguish between three different classes of plants in an agricultural application. In contrast to most of the aforementioned work that relies on clean, standardized top-down images of single leaves or plants, Haug et al. demonstrated their system in a realistic setting, classifying complete, and sometimes overlapping, plants in the field. Unfortunately the authors evaluated on an in-house dataset that is not publicly available.

All of the aforementioned work use manually engineered features. However, a recent trend in the field of object recognition has been to learn features using deep convolutional networks (ConvNets). The most prominent example of this trend is the annual *ImageNet Large Scale Visual Recognition Challenge* where for the past two years many of the participants have used ConvNet features [23].

Convolutional networks themselves are not a new approach and were proposed by LeCun et al. in 1989 [16] to recognize hand-written digits. However, their popularity has risen recently due to algorithmic improvements such as dropout and so called rectified linear units [9, 3] and the availability of computational resources (including GPUs to perform parallel computation) that are needed to train these models. Several research groups have shown that ConvNets outperform more classical approaches for object classification or detection that are based on hand-crafted features [14, 24, 5, 6, 22].

Despite the proclaimed success of the above mentioned features, leaf classification has often been applied to highly controlled images such as those shown in Figure 2. In this work, we aim to test the robustness of some hand-crafted and ConvNet features for leaf classification under varying conditions. This will be done systematically introducing variations to the leaf images. These variations will simulate some of the conditions which are encountered when the imaging conditions are not well controlled.

3. Systems Examined

We analyse a range of features and how well they perform when condition variations are introduced. The features include: ConvNet features, HoCS, and HCFs commonly used within the literature for leaf classification. To ensure the focus of this analysis is on the features being used and not the classifiers, we keep the classifier constant by always using a random forest classifier. While improvements in accuracy may be achieved using other classifiers, this shall not be examined within this work. We also investigate the effectiveness of combining ConvNet features with the more commonly used HCFs.

3.1. Deep Convolutional Neural Network (ConvNet)

To analyse the effectiveness of ConvNet features, we utilise the pre-trained ConvNet framework called Caffe [10]. We follow the idea of Razavian et al. [22], who showed that combining the features from the ConvNet with simple classifiers is highly competitive or even superior to classical approaches in a variety of recognition and detection benchmarks. The network provided by Caffe is trained using 1.2 million images and comprises of five convolutional, three max pooling and two fully connected layers from which potential features can be obtained. Empirically, we found that the last two fully connected layers (layers fc6 and fc7) performed best with little difference between them. Given this, we chose to use layer fc7 for our experiments and the classification system will be referred to as **ConvNet**.

3.2. Histograms of Curvature over Scale (HoCS)

HoCS features used for our analysis are extracted using the same method as described in [15]. This comprises of using discs at 25 different scales around the contour of the leaf in order to describe the shape of the leaf by measuring the percentage of the discs occupied by leaf pixels. As [15] gives no indication as to which scales were used, we use discs with radius (r) values of $r = [5, 6, 7, \ldots, 30]$ pixels[2]. The classification system using HoCS features will be referred to as **HoCS**.

3.3. Hand Crafted Features (HCFs)

In order to analyse more commonly used hand crafted features for plant classification, the shape and statistical features described in [8] were used. These features are summarized in Table 1. The classification system using all of these features will be referred to as **HCF**. In [8] the statistical features were calculated using the NDVI vegetative index image, we approximate this by using the normalized excessive green (NExG) vegetative index image [7]. This is done as the images used within the experiments do not contain the near infrared (NIR) information required to generate a NDVI vegetative index image.

[2]Source code can be found at `http://tinyurl.com/AGR-QUT`

Feature	Scale Robust
Perimeter (length of contour)	No
Area (number of pixels covered by leaf)	No
length of skeleton	No
Compactness (area / perimeter^2)	Yes
Solidity (area / area of convex hull)	Yes
Convexity (perimeter)	Yes
length of skeleton/perimeter	Yes
minimum of leaf pixel intensities	Yes
maximum of leaf pixel intensities	Yes
range of leaf pixel intensities	Yes
mean of leaf pixel intensities	Yes
median of leaf pixel intensities	Yes
standard deviation of leaf pixel intensities	Yes
kurtosis of leaf pixel intensities	Yes
skewness of leaf pixel intensities	Yes

Table 1: Hand-crafted shape and statistical features used for analysis, as well as whether feature was considered scale robust or not.

We note that many shape and statistical features are inherently robust to scale. As such, we create another HCF classification system known as **HCF-ScaleRobust**. This uses the features marked as scale robust in Table 1.

To test if some classification benefit can be gained through combining ConvNet features and commonly used HCFs, we propose a new classification system **Combined**. Further details as to how this system is used for classification are given in Section 3.4. We believe this system will be particularly interesting in the presence of condition variations as we believe that the ConvNet features, which have not seen all of these potential variations, do not have the innate robustness that the hand-crafted features have.

3.4. Classifiers

To compare the performance of different feature types, we use a random forest classifier for all the features examined. Random forests are a technique which was introduced in [1], and consists of generating a large quantity of decision trees. Decision trees are a series of boolean statements regarding values of certain features which give a classification based on the results of said boolean statements [1].

We optimise the parameters of the random forest by choosing the maximum depth and number of trees which optimises performance on the validation set, see Section 4 for details on the validation set. The possible values for maximum depth (D) were $D = [5, 10, 15]$, as well as the option for there being no maximum depth. The possible number of trees (T) in the random forests is $T = [25, 50, 75]$. This was done for **ConvNet**, **HoCS**, **HCF** and

HCF-ScaleRobust.

We combine the ConvNet features with the HCFs to form a **Combined** classifier. This is achieved by combining the probability scores given by both **ConvNet** and **HCF-ScaleRobust** via the sum rule [13]. The class which has the highest score after this summation is then taken to be the best match.

4. Experimental Results

We conduct experiments using the Flavia dataset [26]. This consists of $1,907$ leaf images of 32 species with at least 50 images per species and at most 77 images. Experiments are performed using a protocol which ensures that the leaf classification systems are optimised on data that is independent of the final evaluation set. Using the original images, we determine a baseline accuracy for each feature type described in Section 3. We then perform tests on modified versions of the images to simulate various real-word conditions. This allows us to evaluate the robustness of the different features to condition variations such as translation, scaling, rotation, shading, and occlusion. In order to allow for translations and rotations of the leaves without losing parts of the leaf, each leaf image from the Flavia dataset was given a white border such that it doubled the image's size.

When extracting features from the images, both the HoCS and the hand-crafted shape features require a segmentation step in order to accurately describe the shape of the leaves. To ensure that the experiments focus more on the effectiveness of the different feature types and not on the accuracy of the segmentation, perfect segmentation is assumed to be true. This segmentation is performed using the segmentation method designed to be used with this dataset which is described in [26].

4.1. Evaluation Protocol

To test the classification systems, we define 10 random splits of the data[3]. Each split contains a *train* set, *validation* set and *evaluation* set. The train set is used to derive the classifier whose hyper parameters are optimised on the validation set. Since there are 10 splits, we take the best average performance over the 10 splits. Using the trained classifier with optimised parameters, the final system performance is then obtained on the evaluation set. The validation and evaluation sets both have 10 samples for each species of leaf, the remaining images are allocated to the train set. When training the classifiers, we make use of the original images only to properly evaluate the effect of the condition variations.

As a performance metric, we use the rank-1 identification rate ($Accuracy$). This measure reflects the rate at which the species of interest is the best match (rank-N identification refers to how often the species of interest is in the N best matches) and is given by $Accuracy = \frac{N_c}{N_t} \times 100\%$ where N_c is the number of correct matches (the correct species was the best match), N_t is the total number of test samples, and $Accuracy$ is given in terms of percentage. This accuracy is attained for each of the ten *evaluation* sets and then the average of these ten accuracies ($\bar{A}$) is used for evaluation and comparison of different classification systems.

Standard deviation (σ) was also calculated across the 10 *validation* and *evaluation* sets. This is calculated by $\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(A_i - \bar{A})^2}$ where A_i is the accuracy of the i^{th} split. When displaying results visually, standard deviations are not shown on graphs to reduce clutter.

4.2. Baseline Results

Baseline classification accuracy for the different features was obtained using the original set of images from the Flavia dataset. The results, in Table 2, show that the ConvNet features provide a considerable performance improvement over the previously proposed features. It can be seen that the HoCS features provide the worst performance with an absolute difference of 17.3% when compared to **HCF-ScaleRobust**. The results of **HCF** and **HCF-ScaleRobust** are also seen to be slightly worse than the best system from the original Flavia dataset tests which achieved 94% classification accuracy [26]. Finally, the use of the **Combined** system, produced the best baseline classification accuracy.

	Valid. Accuracy	Eval. Accuracy
Combined	$\mathbf{96.6\% \pm 1.3\%}$	$\mathbf{97.3\% \pm 0.6\%}$
ConvNet	$95.4\% \pm 1.0\%$	$94.5\% \pm 1.1\%$
HCF	$90.5\% \pm 1.1\%$	$91.2\% \pm 1.6\%$
HCF-ScaleRobust	$89.3\% \pm 1.6\%$	$89.8\% \pm 1.6\%$
HoCS	$72.0\% \pm 2.2\%$	$70.2\% \pm 1.7\%$

Table 2: The baseline rank-1 identification rate (accuracy) for five systems including: **Combined**, **ConvNet**, **HCF**, **HCF-ScaleRobust**, and **HoCS**. In bold is the best performing system.

4.3. Condition Variations

To test the robustness of each classification system described in Section 3, we simulate real-world conditions by introducing a set of condition variations to the data. These condition variations are introduced in a structured manner to ensure we can measure the impact of each condition. We examine five different forms of condition variations: translations, scaling, rotations, shading and occlusions. To the

[3]Our protocols can be found at `http://tinyurl.com/AGR-QUT`

(a) (b) (c)

Figure 3: Simulated real-world condition on original Flavia dataset images: (a) original leaf image, (b) shaded leaf image with 40% average V decrease, (c) occlusion leaf image with a 5% occlusion.

best of our knowledge, this is the first time that such a detailed analysis has been conducted for leaf classification.

4.3.1 Translation Analysis

In the Flavia dataset our classifier is trained on, the leaves typically appear centered in the images. However, when capturing images in real-world scenarios, the precise position of a leaf will not be stable. To simulate this variability in positioning, we translate the leaves of the evaluation images horizontally (t_h) and vertically (t_v) independently. The translations, t_h and t_v, are applied as a percentage of the original image size (before the white borders are added) to a maximum of 50% that $t_h = [10, 20, \ldots, 50]$ and $t_v = [10, 20, \ldots, 50]$.

Since **HCF**, **HoCS**, and **HCF-ScaleRobust** are calculated on an already segmented leaf, they are totally invariant to the original position of that leaf in the image. The performance of these features therefore remains stable at their baseline performance. The translation test was mainly conducted to measure the stability of the **ConvNet** system, which is not inherently invariant to translations of object of interest within its input image. However, as can be seen in Figure 4, **ConvNet** remains fairly stable until 30% translation is reached and its performance degrades below the accuracy of **HCF** and **HCF-ScaleRobust**. The initial stability is provided by the pooling layers contained within the ConvNet architecture. Extra robustness to translational variation is achieved by **Combined** which combines **ConvNet** and **HCF-ScaleRobust** and remains the best classification system throughout.

4.3.2 Scaling Analysis

The scale of the observed leaves can vary significantly between the training dataset and the images acquired during deployment. We therefore want to evaluate the influence of scale changes between training and test data and scale the leaves within the evaluation images to a percentage $s = [90, 80, \ldots, 40]$ of the original image size.

After testing, we found that, as expected, the scale robust subset of the hand-crafted features were robust to when the scale of the leaf image is altered as this was the purpose for

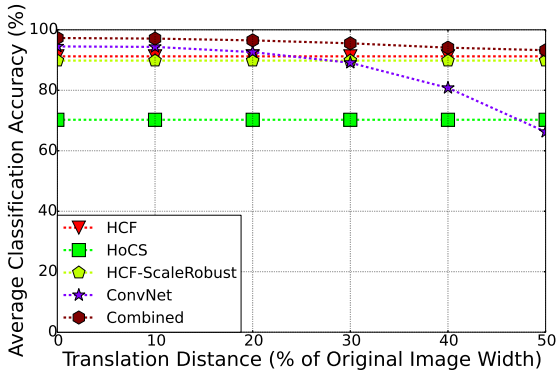

Figure 4: Average classification accuracy across ten testing sets for **HCF**, **HoCS**, **HCF-ScaleRobust**, **ConvNet** and **Combined** systems when subjected to horizontal translation

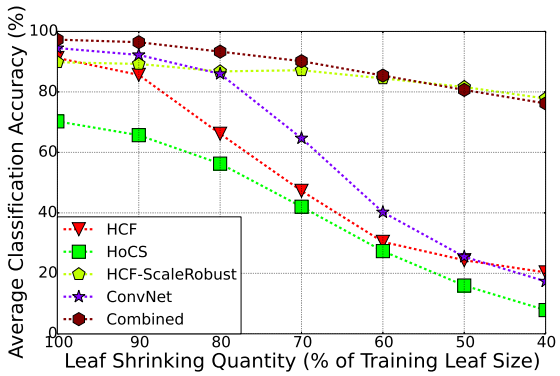

Figure 5: Average classification accuracy across ten testing sets for **HCF**, **HoCS**, **HCF-ScaleRobust**, **ConvNet** and **Combined** systems when subjected to scaling

which they were designed. In comparison to this, however, the ConvNet, HoCS, and hand-crafted features were not.

In Figure 5, it can be seen that the performance of the **ConvNet**, **HCF** and **HoCS** systems degrades considerably when the scale changes. The difference in robustness between **HCF** and **HCF-ScaleRobust** shows the significant negative impact to scaling robustness that the three non-scale robust features within **HCF** have. **HCF** is shown to have lower classification accuracy than **HCF-ScaleRobust** after only a 10% decrease in leaf size.

Similar to the translation test above, we observed **ConvNet** to remain relatively stable for scale changes up to 20% and this can be explained by the invariance introduced by the pooling layers. Despite the considerable degradation in performance of **ConvNet** for larger scale changes, the **Combined** system still has stable performance and provides the best overall accuracy up to a 50% decrease in leaf size.

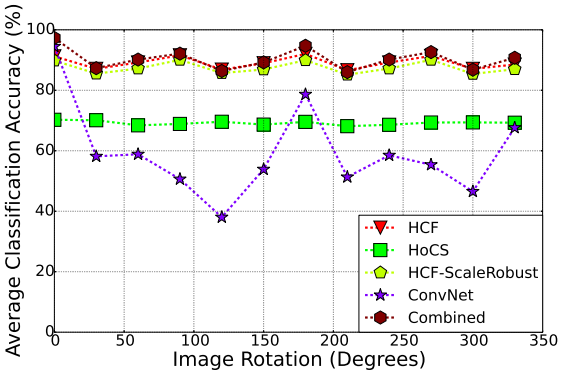

Figure 6: Average classification accuracy across ten testing sets for **HCF**, **HoCS**, **HCF-ScaleRobust**, **ConvNet** and **Combined** systems when subjected to rotation

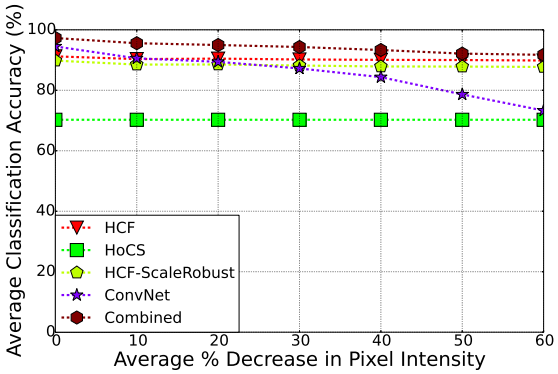

Figure 7: Average classification accuracy across ten testing sets for **HCF**, **HoCS**, **HCF-ScaleRobust**, **ConvNet** and **Combined** systems when subjected to shading

4.3.3 Rotation Analysis

Rotation is simulated as the orientation of a leaf cannot be guaranteed in real-world scenarios. To examine the impact of rotation we use a coarse set of rotations consisting of $30°$ rotations such that the set of rotations examined is $d = [30°, 60°, 90°, \ldots, 330°]$.

From the results of the classification tests on these rotated images, which are summarized in Figure 6, it can be seen that the hand-crafted and HoCS features are robust to rotations. As with translation, this is because they are designed to work using a segmented image that will not change with rotation. By contrast, the ConvNet features are not robust to rotations with performance dropping as low as 38%. ConvNets are not designed to be rotation invariant and therefore fail under this condition. Despite this instability, the highest performance over most rotation levels is given by the **Combined** system. Also it can be seen that **HoCS**, while stable, performs considerably worse than **Combined**, **HCF** or **HCF-ScaleRobust**.

Finally, it is noticed there is cyclic performance degradation between $0°$, $90°$, $180°$, and $270°$. We attribute this to the fact the when rotating the images between these points interpolation is required leading to minor degradation in performance for all of the systems.

4.3.4 Shading Analysis

While images for the training set can be collected under controlled conditions, it is much harder to control the illumination conditions in real world applications. To examine the potential effect of different illumination conditions between training and test set, an artificial shading effect was applied to the leaf pixels within each image. To produce this shading effect, we decrease the intensity value (V) of each of the leaf pixels in HSV color space by a normally distributed random percentage value with a standard deviation of 10%. For the tests we vary the mean percentage values $v_\mu = [10, 20, \ldots, 60]$.

Figure 7 shows that the various features coped well with the induced changes in illumination. The statistical features within **HCF** are calculated on normalized colour values and are therefore not affected by a simple uniform change in illumination brightness. We expect to see worse results when inducing non-uniform changes across different color changes, that might be caused by a non-perfect white balance during testing. A purely shape feature as **HoCS** is of course not affected by shading, as long as the necessary pre-segmentation is still perfect. Since the features used by **ConvNet** are calculated on the raw pixel color values, they are not invariant to changes in illumination and we can observe a slight degradation of performance in Figure 7 .

Again, **Combined** was seen to have the highest, and **HoCS** the lowest, classification accuracy across all tests.

It should be noted, that while all feature systems were seen to cope fairly well with this uniform illumination change, we would expect a different result if localized shading were to occur. We leave this open for future work but surmise that it would likely detrimentally affect the statistical HCFs and cause a decrease in accuracy for the **HCF**, **HCF-ScaleRobust** and **Combined** systems.

4.3.5 Occlusion Analysis

Occlusions of leaves can occur for many reasons including damage to the leaf or, occlusion by other leaves of the same or different neighbouring plants. To generate such occlusion effects, we remove a part of the leaf in the image by artificially overlaying a white disk of varying diameter onto the leaf. Given that a leaf within an image has area A_{leaf} we place the centre of the disk randomly on the leaf. The area of the disk is chosen to be $p = [5, 10, 15, \ldots, 30]$ percent of A_{leaf}. An example of this process is given in Figure 3c.

From Figure 8 it can be seen that the none of the tested features are robust to occlusion. However, the ConvNet fea-

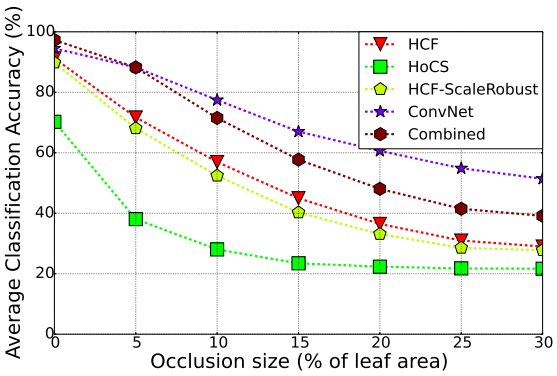

Figure 8: Average classification accuracy across ten testing sets for **HCF**, **HoCS**, **HCF-ScaleRobust**, **ConvNet** and **Combined** systems when subjected to occlusions

tures are least degraded and exhibit the best overall performance. The **Combined** system provides consistent performance which is worse than the **ConvNet** system but considerably better than the **HCF**, **HoCS** or **HCF-ScaleRobust** systems. It can also be seen that again, **HoCS** provides the lowest classification accuracy over all tests.

4.4. Summary of Results

Several observations can be made as to the robustness of different classification systems for the purpose of leaf classification.

- **ConvNet** gives better baseline accuracy than **HCF**, **HCF-ScaleRobust** and **HoCS**.
- **Combined** shows greater robustness than **ConvNet** for all tests except occlusion tests.
- None of the tested features are robust to occlusions, but **ConvNet** copes best with such disturbances.
- Removing the three scale variant features within **HCF** gives vast improvement to scaling robustness without major loss in overall accuracy in other circumstances.
- **ConvNet** has low robustness to translation and shading and no robustness to rotations.
- **Combined** outperforms the best performing HCF system (**HCF-ScaleRobust**) by 5.7% averaged across all tests.
- While stable to all conditions except scaling and occlusion, **HoCS** [15] performs worst out of all classification systems.

From these observations it becomes apparent that the main weakness of the traditional hand crafted features is that they are susceptible to occlusions. Part of the reason that **HCF**, **HCF-ScaleRobust**, **HoCS** and **Combined** performed well under other types of condition variations is because they rely on a perfect segmentation of the leaf from the background. In real-world applications, perfect segmentation is not always possible which will lead to a significant change of the shape properties (as simulated by our occlusion test), but also of the statistical features based on normalized color.

ConvNet on the other hand has exhibited a mild invariance to reasonable shape changes such as occlusions. It also reaches the best baseline performance of all tested single systems. Related work in the computer vision community, especially on the ILSRVC benchmarks [23] has demonstrated that ConvNet-based systems are very well able to perform object detection and classification *without* having to perform a foreground-background segmentation first. We therefore expect ConvNet systems to perform well when trained and tested with natural backgrounds in an application for the real world. Extensive tests of this hypothesis are left for future work. Furthermore, the lacking invariance of ConvNet features against rotation can be overcome by including rotated leafs in the training set. We tested this in an additional experiment and could observe that the rotation influence on the **ConvNet** system was completely removed and it remained stable at its baseline performance.

5. Conclusion and Future work

We have presented an extensive evaluation of features for leaf classification under a variety of condition variations such as translation, scaling, rotation, shading and occlusion.

We empirically demonstrated that combining ConvNet and HCF features leads to a state-of-the-art leaf classification system (**Combined**). For ideal conditions, this **Combined** approach yields state-of-the-art performance with an average accuracy of $97.3\% \pm 0.6\%$ compared to traditional features which obtain an average accuracy of $91.2\% \pm 1.6\%$. Furthermore, for the range of condition variations examined, this system consistently outperforms the best HCF system (**HCF-ScaleRobust**) on average by 5.7%.

There are several possible directions for future work. First we want to explore how ConvNet features perform on images of leaves and complete plants taken under real-world conditions with natural background. This requires training the classifier using features from a variety of images collected under application-like conditions, e.g. for agricultural applications. Creating such a large realistic training dataset would also help the community and foster future work on plant classification in the wild. Since our tests showed that combining **ConvNet** and **HCF-ScaleRobust** into the **Combined** system led to the best overall performance, future work should also explore how different complementary features can be combined in a more beneficial way than proposed here.

Acknowledgements

This work has been supported by the Department of Agriculture Fisheries and Forestry (DAFF) of the Queensland government through the Strategic Investment in Farm Robotics (SIFR) at QUT.

References

[1] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[2] S. Cho, D. Lee, and J. Jeong. Automation and emerging technologies: Weed–plant discrimination by machine vision and artificial neural network. *Biosystems Engineering*, 83(3):275–280, 2002.

[3] G. E. Dahl, T. N. Sainath, and G. E. Hinton. Improving deep neural networks for LVCSR using rectified linear units and dropout. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 8609–8613. IEEE, 2013.

[4] F.-M. De Rainville, A. Durand, F.-A. Fortin, K. Tanguy, X. Maldague, B. Panneton, and M.-J. Simard. Bayesian classification and unsupervised learning for isolating weeds in row crops. *Pattern Analysis and Applications*, pages 1–14, 2012.

[5] 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. *arXiv preprint arXiv:1310.1531*, 2013.

[6] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *arXiv preprint arXiv:1311.2524*, 2013.

[7] J. M. Guerrero, G. Pajares, M. Montalvo, J. Romeo, and M. Guijarro. Support vector machines for crop/weeds identification in maize fields. *Expert Systems with Applications*, 39(12):11149–11155, 2012.

[8] S. Haug, A. Michaels, P. Biber, and J. Ostermann. Plant classification system for crop/weed discrimination without segmentation. In *IEEE Winter Conference on Applications of Computer Vision*, 2014.

[9] 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.

[10] Y. Jia. Caffe: An open source convolutional architecture for fast feature embedding. http://caffe.berkeleyvision.org, 2013.

[11] A. Joly, H. Müller, H. Goëau, H. Glotin, C. Spampinato, A. Rauber, P. Bonnet, W.-P. Vellinga, and B. Fisher. Lifeclef 2014: multimedia life species identification challenges. In *Proceedings of CLEF 2014*, 2014.

[12] S. Kiani. Discriminating the corn plants from the weeds by using artificial neural networks. *International Journal of Natural & Engineering Sciences*, 6(3), 2012.

[13] J. Kittler, M. Hatef, R. Duin, and J. Matas. On Combining Classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:226–239, 1998.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. Burges, L. Bottou, and K. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[15] 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*, pages 502–516. Springer, 2012.

[16] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.

[17] W. S. Lee, D. Slaughter, and D. Giles. Robotic weed control system for tomatoes. *Precision Agriculture*, 1(1):95–113, 1999.

[18] C. Lin. *A support vector machine embedded weed identification system*. PhD thesis, University of Illinois, 2009.

[19] S. Mouine, I. Yahiaoui, and A. Verroust-Blondet. A shape-based approach for leaf classification using multiscaletriangular representation. In *Proceedings of the 3rd ACM Conference on International Conference on Multimedia Retrieval*, ICMR '13, pages 127–134, New York, NY, USA, 2013. ACM.

[20] H. Okamoto, T. Murata, T. Kataoka, and S.-I. HATA. Plant classification for weed detection using hyperspectral imaging with wavelet analysis. *Weed Biology and Management*, 7(1):31–37, 2007.

[21] A. Piron, V. Leemans, F. Lebeau, and M.-F. Destain. Improving in-row weed detection in multispectral stereoscopic images. *Computers and electronics in agriculture*, 69(1):73–79, 2009.

[22] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson. CNN features off-the-shelf: an astounding baseline for recognition. *arXiv preprint arXiv:1403.6382*, 2014.

[23] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. Imagenet large scale visual recognition challenge. *arXiv preprint arXiv:1409.0575*, 2014.

[24] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229*, 2013.

[25] D. Slaughter, D. Giles, and D. Downey. Autonomous robotic weed control systems: A review. *Computers and electronics in agriculture*, 61(1):63–78, 2008.

[26] S. G. Wu, F. S. Bao, E. Y. Xu, Y.-X. Wang, Y.-F. Chang, and Q.-L. Xiang. A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network. In *IEEE International Symposium on Signal Processing and Information Technology*, pages 1–6, 2007.