

Image identification of *Protea* species with attributes and subgenus scaling

Peter Thompson
Stellenbosch University
peter@hoekwil.com

Willie Brink
Stellenbosch University
wbrink@sun.ac.za

Abstract

The flowering plant genus *Protea* is a dominant representative for the biodiversity of the Cape Floristic Region in South Africa, and from a conservation point of view important to monitor. The recent surge in popularity of crowd-sourced wildlife monitoring platforms presents both challenges and opportunities for automatic image based species identification. We consider the problem of identifying the *Protea* species in a given image with additional (but optional) attributes linked to the observation, such as location and date. We collect training and test data from a crowd-sourced platform, and find that the *Protea* identification problem is exacerbated by considerable inter-class similarity, data scarcity, class imbalance, as well as large variations in image quality, composition and background. Our proposed solution consists of three parts. The first part incorporates a variant of multi-region attention into a pre-trained convolutional neural network, to focus on the flowerhead in the image. The second part performs coarser-grained classification on subgenera (superclasses) and then rescales the output of the first part. The third part conditions a probabilistic model on the additional attributes associated with the observation. We perform an ablation study on the proposed model and its constituents, and find that all three components together outperform our baselines and all other variants quite significantly.

1. Introduction

The iconic plant genus *Protea* has its centre of diversity in the Cape Floristic Region (CFR) of South Africa; a region that accounts for 40% of the country's 20,400 species of indigenous flowering plants [30] while covering only 4% of the country's area. The diversity of *Protea* makes it a fitting surrogate for the biodiversity of the region [8] and consequently an important genus to monitor for the sake of conservation.

The monitoring of biodiversity is traditionally performed by expert scientists, but there is a growing trend to utilise the power of crowd-sourced data [5, 26]. Such data is becoming



Figure 1. Different species of *Protea*, such as *Protea neriifolia* and *Protea laurifolia* shown here, can exhibit considerable visual similarity.

important for understanding species populations [3] in the midst of issues like global warming, pollution and poaching. The crowd-sourced platform iNaturalist for example allows users to upload observations of wildlife, which typically include images, locations, dates, and identifications that can be verified by fellow users [31]. As of October 2019 the iNaturalist database contains over 27,000,000 observations for over 237,000 species, and it is impossible for experts to keep up with the sheer influx of data [33].

Automated tools based on computer vision may ease the task of identification, and could potentially provide expert-like knowledge to amateur naturalists. iNaturalist implements a top- k recommender system built on deep convolutional models for image identification [31], but challenges due to large class imbalances and fine granularity in biological domains remain [34, 3].

We focus on the problem of automatically identifying *Protea* species from images, as a surrogate both for the biodiversity of the CFR and for the unbalanced and fine-grained databases of citizen science projects in general. The problem is complicated by a number of factors. Firstly, it is a fine-grained classification problem where some species share striking visual similarities with others, as demonstrated in Figure 1. Secondly, image data is extremely scarce for many of the rarer species. When we constructed

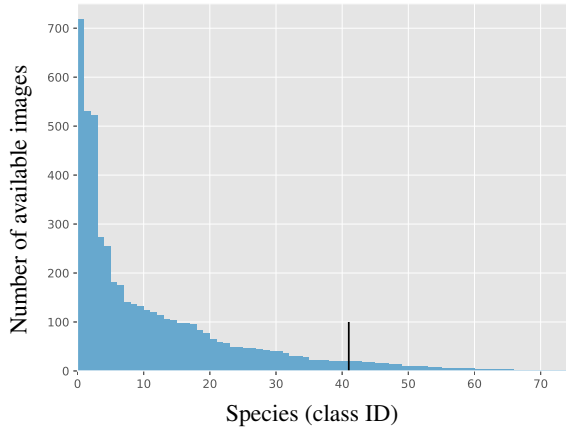


Figure 2. The distribution of images per species in our dataset, indicating a degree of class imbalance and long tail. Species left of the short vertical line are those for which at least 20 images are available.

our dataset (as detailed in section 3), only 41 of the 70 *Protea* species known to exist in the CFR had at least 20 different images depicting an inflorescence (flowerhead). Unfortunately the prevalence of hybrid cultivars prevent scraping the Internet for additional images, labelled or otherwise. Thirdly, the data is unbalanced as indicated in Figure 2. Four of the 41 species mentioned above account for nearly 40% of the data. Finally, the image data is sourced from populations in the wild, by many different observers. There is no standard in how images were taken, resulting in large amounts of compositional and background variation.

In order to address these challenges we restrict the problem to the 41 species for which at least 20 images could be found, and propose an automated identification model that consists of three components. The first is a convolutional neural network (CNN) with a variant of multi-region attention [38], trained to classify over the 41 species. The second component leverages the fact that *Protea* species can be categorised into more easily distinguishable subgenera (the two species in Figure 1 are both Bearded Sugarbushes, for example) and accordingly consists of a CNN trained for subgenus classification. Its output is used essentially to rescale the class scores of the first network. The occurrence of *Protea* species tends to be relatively finely dependent on location and elevation, and different species also flower during different times of the year. Such attributes are often available as part of an observation, and the third component of our model exploits such additional data (if available) through a simple Bayesian approach.

An ablation study on the proposed model suggests that all three components together outperform the baselines substantially in terms of test accuracy and recall. The image dataset can be found within a project called “Sugarbushes of South Africa” on iNaturalist.

2. Related work

The idea of image recognition for automated species identification has been around for some time [11]. While a lot of previous work on plant identification rely on hand-crafted features, recent advances in deep feature learning are opening new opportunities for more challenging, fine-grained identification.

2.1. Feature engineering

A review on plant species identification by Wäldchen *et al.* [32] mentions that different plant organs such as leaves, flowers, fruit and stems, have typically been considered. It is then common to define and extract shape, colour or texture features from these organs [7, 14, 15, 19, 20, 21]; a trend that has seen some continuation beyond the deep learning revolution [1, 6, 22, 23, 36].

Hong and Choi [13], for example, identify flowers based on detected edge contours and also colour features found through clustering in the HSV space. Aprianti *et al.* [1] identify orchid species from images by first segmenting flowers from the background and then finding shape and colour features. Shape features include segment size, aspect ratio and roundness, while the saturation component of the segmented image is used as a colour feature. Zawbaa *et al.* [36] use textures, by transforming an input image to a set of binary images and extracting texture patterns. Nilsback and Zisserman [19] describe texture by convolutional filters.

2.2. Deep learning based approaches

Recent work on plant identification tend to make use of deep neural networks that learn statistically relevant feature representations from the data. Zhang *et al.* [37] train a six-layer CNN for leaf identification on the Flavia dataset [35]. Barré *et al.* [2] find improved performance on the same task with a 17-layer CNN, and also show significant improvements over the use of hand-crafted features.

A common approach is to leverage the power of pre-trained CNNs such as AlexNet [17] or ResNet [12]. Simon and Rodner [27] use this idea as a method of feature extraction, with an SVM for classification, on the Oxford Flowers 102 dataset [20]. Given the challenges in our particular case, notably the data scarcity, we also utilise a pretrained network as a basis in our approach.

There have also been advances in applying active learning to fine-grained image classification [16], which seeks to increase the training set on-the-fly in a manner that optimises network performance. It is not really applicable to our niche problem, where *Protea* species are sometimes rare, not well documented, or found only in difficult to reach environments.

2.3. Fine-grained image recognition

A standard approach to tackle fine-grained recognition is to localise and then separately classify diagnostic features in images with a number of subnetworks. Zheng *et al.* [38] use the outputs of a pretrained CNN to construct an attention mechanism for a subsequent classification subnetwork. This technique does not rely on part-based annotations of training images (as some earlier methods did), and has become a norm in fine-grained recognition [10, 28] and one that we also adopt.

Another popular method for fine-grained classification is end-to-end feature encoding with bilinear CNNs [18]. Such a model consists of two parallel CNNs acting as feature extractors, whose outputs are multiplied and pooled to obtain an image descriptor vector for classification.

2.4. Incorporating additional data

Additional data, such as the recorded location of an observation, has been used to solve biological classification problems. An example of this is BirdSnap [4] which, through an application of Bayes' rule, takes geographical distributions of bird species into consideration to aid an image based identification module. We incorporate location, elevation and date information in a similar way.

3. Dataset

This section describes our process of collecting an annotated dataset of images of *Protea* species, as well as our construction of per-species distributions according to location, elevation and time of flowering.

3.1. iNaturalist

We collected images from the crowd-sourced platform iNaturalist, where people across the world upload observations of fauna and flora in the wild, under a Creative Commons license. An observation typically consists of at least an image, a location, a date and a community-aided identification. Of all the *Protea* records found on iNaturalist at the time of our dataset creation, we were interested only in those from non-cultivated, research-grade (having two or more unanimous species-level identifications from separate users) observations in the CFR. We also kept only images depicting flowering inflorescences, and restricted the dataset to species with at least 20 such images. This filter process resulted in a dataset containing 4,849 images in total, across 41 species.

We emphasise that the set is unbalanced in terms of samples per species, has fine granularity among many of the classes, and also contains significant variability in background, image quality, image composition and the size of the inflorescence in the image, etc. It is, however, representative of the real world [31].

Every image corresponds to a latitude and longitude value of where it was taken, an elevation reading in metres above sea level, the date of the observation, and a community identification to species level. We note that elevation can to some degree be inferred from latitude and longitude, but we rather treat it as an additional attribute because of the sensitivity of certain *Protea* species to it. We also include the iNaturalist observation identification number in our dataset, for potential future use (to trace a specific observation, for example).

We split our dataset into a training set with 3,652 images and a test set with 1,197 images, by splitting the images of each of the 41 classes randomly with a fixed ratio.

3.2. The Protea Atlas Project

The Protea Atlas Project [24] was launched in November 1991 by Rebelo, in an effort to document the *Proteaceae* in Southern Africa. The project culminated in a vast collection of data: 252,513 species records at 61,591 locations.

We isolate the data for our 41 *Protea* species, for an indication of where each species is found. We discretise the CFR into a gridmap, and for each species separately populate the grid cells with frequency counts from the Protea Atlas Project records. These frequencies are normalised and then interpreted as a conditional probability distribution over observation location, given a species. An example of such a distribution is shown in Figure 3.

We construct similar distributions over elevation and flowering time, using the summarised data in Rebelo's field guide [25]. For every species we set up binary-valued distributions over discrete elevation intervals (in steps of 100m) and discretised flowering time of year (in months). These values are then smoothed with a simple 1D Gaussian filter to reduce potential quantisation effects, and normalised.

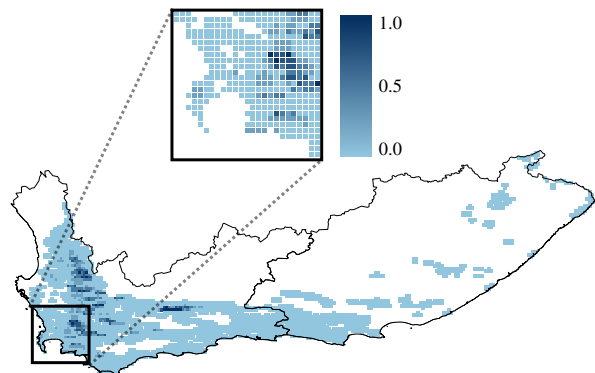


Figure 3. The location distribution of *Protea magnifica*, as inferred from the Protea Atlas Project records. Every cell in the discretised Cape Floristic Region is shaded according to the probability of occurrence in that location, given the species.

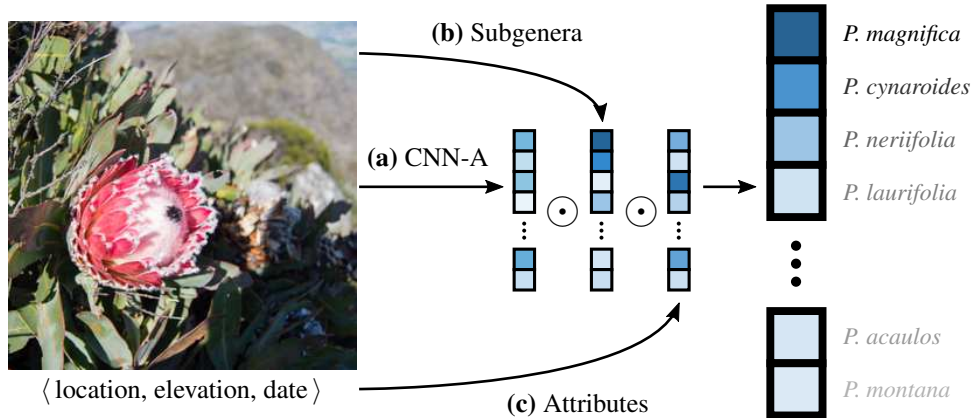


Figure 4. Our model for *Protea* species identification operates on an image with additional (but optional) attributes linked to the observation, and combines three parts: (a) a CNN with attention, (b) a separate network that classifies the image into coarser subgenera, and (c) a probabilistic model conditioned on the attributes.

4. Our approach

An overview of our approach is given in Figure 4. The goal is to perform *Protea* species identification from an observation, which we assume consists of a single image and additional (but optional) position and date information. The proposed model has three parts, each to be explained in more detail in the subsections below.

The first part is a convolutional neural network with attention (CNN-A) that outputs a normalised score for each of the 41 possible species (classes). The second part leverages the fact that the 41 species can be grouped into 13 subgenera (superclasses) that are more distinct from one another. This somewhat easier classification problem is solved with a CNN, whose output is used essentially to rescale the class scores from the first part. The third part of our model conditions the distribution over species on evidence of location, elevation and date of the observation. The final result from the three parts is a classification score vector.

4.1. CNN with attention

The first part of our model makes use of a CNN to transform images to normalised class scores associated with the different species. As a first baseline we make use of the Inception-V3 architecture [29] with weights pretrained on ImageNet [9]. We choose this particular architecture for its good balance between complexity and performance. We freeze the convolutional layers, replace the last five fully-connected layers such that a 41-class softmax output is produced, and train the network on our data.

Prompted by the unconstrained nature of images from field observations, as well as the potentially large variations in backgrounds, we opt to explore the inclusion of an attention mechanism. We base this component on the multi-

region method of Zheng *et al.* [38], which learns to find a preset number of attention regions in an image specifically for fine-grained classification. We extract two regions per image: one that ought to focus on the prominent inflorescence, and one that might pick up salient areas in the background. Only the first of these is passed to the next phase of the model. Through informal experiments we found this approach to perform better than a single-region attention model, likely because of the extra constraints that the second region imposes on the first during training.

More specifically, a 299×299 colour image is fed into the convolutional base of a pretrained Inception-V3 network, yielding 2,048 feature maps each of size 8×8 . The idea now is to create two separate combinations of these feature maps that will form the two attention maps over the given image. The transformation from feature maps to attention maps can be performed by two fully-connected neural networks [38]. The feature maps are first clustered into two groups by k -means on their peak responses over the training set, and then averaged per cluster into attention maps M_1 and M_2 . The networks are initially trained to reproduce these maps, and then further fine-tuned under a grouping loss that favours tightness within each map and dissimilarity between them. This loss is computed over attention map i , where $i \in \{1, 2\}$ and may be expressed as

$$L_{\text{group}}^{(i)} = \sum_{(x,y)} M_i(x,y) [(x - p_x)^2 + (y - p_y)^2] + \lambda \sum_{(x,y)} M_i(x,y) [M_{3-i} - \alpha], \quad (1)$$

where $M_i(x,y)$ is the value of attention map i at grid location (x,y) . Coordinates (p_x, p_y) represent the location of the maximum value of M_i , and α is a scalar margin. The

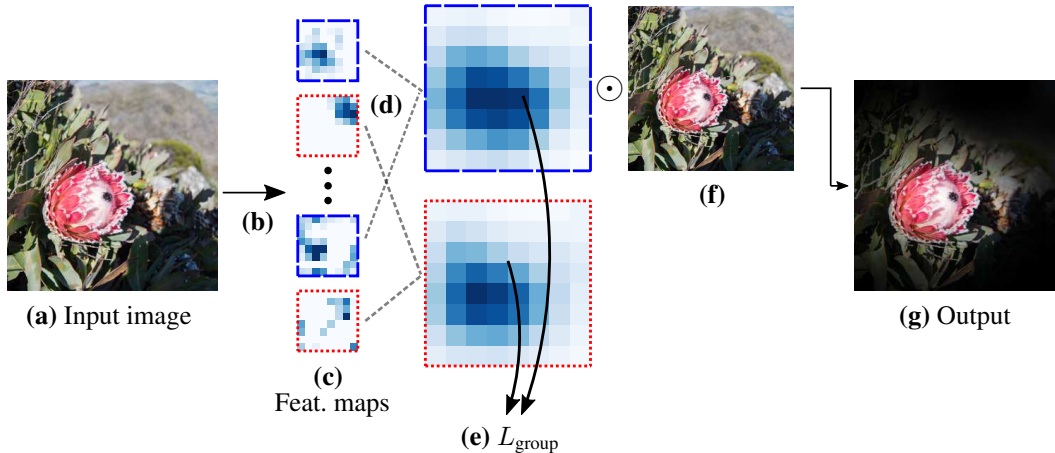


Figure 5. An image (a) is passed through the convolutional layers of Inception-V3 (b), which leads to 2,048 feature maps (c). These are combined into attention maps through two fully-connected networks (d), learned jointly through the minimisation of a group loss (e). The most prominent of the two maps is scaled and multiplied with the original image (f), to produce an attention-boosted image (g).

importance of the first term (for in-map tightness) relative to the second term (for between-map dissimilarity) is controlled by the hyper-parameter λ .

As mentioned above we take only one of the two attention maps further; ideally the one that focuses on the *Protea* inflorescence. Through some experimentation on our training set we found that this map is easily discernible as being by far the largest of the two clusters from k -means (which again can be attributed to the nature of the data, where almost all images contain the inflorescence as a large salient object). The attention map is upscaled and element-wise multiplied with the input image, as illustrated in Figure 5. The resulting attention-boosted images are used to train a CNN similar to the fine-tuned Inception-V3 network described at the beginning of this section.

The attention map extractor and the attention-boosted image classifier can be optimised end-to-end, or alternately for a number of iterations (similar to what is done in [38]).

4.2. Subgenus scaling

The 41 species of *Protea* in our dataset can be grouped into 14 subgenera, according to common traits, and the second part of our model attempts to classify a given image into one of these subgenera. It can be regarded slightly easier than the 41-class problem above, due to the 14 classes being less fine-grained and more distinct, the data being less unbalanced, and the availability of more samples per class. We employ a pretrained Inception-V3 network, replace the last five fully-connected layers, impose a 14-class softmax layer as output, and train the new layers with our data. Note that relabelling our training data from species to subgenera is straightforward with a guide like [25]. We experimented with an attention mechanism in this network as well, but

found no significant change in performance. It might be due to the simplified nature of the problem, which already leads to a marked improvement in accuracy (as we see in section 5).

The subgenus classifier produces 14 class scores for a given image, which we transform into scores over the 41 *Protea* species by distributing the score of each parent subgenus equally among its children. Here we essentially assume a uniform distribution over the species given the subgenus. An alternative would be to incorporate the class imbalance over the species, but there is a risk of overcompensation since the species-level CNN with which the subgenus classifier is to be combined might already be learning the class imbalance. The 41 scores produced here are used to rescale the output of the CNN-A model from section 4.1, through element-wise multiplication.

4.3. Attributes

For the third part of our model we consider the possible availability of three attributes accompanying an image, namely location, elevation and date. The location can be mapped to our discrete grip map from section 3.2. Similarly, the elevation is binned to one of our discrete intervals. Since we consider only observations of *Protea* species in flower, the date can be interpreted as an observation of flowering time.

We combine the three attributes x_1, x_2, x_3 with a simple Bayes model that assumes conditional independence between them:

$$p(y_i | x_1, x_2, x_3) \propto p(y_i) p(x_1 | y_i) p(x_2 | y_i) p(x_3 | y_i), \quad (2)$$

where y_i is the event of the observation being species i . The

conditionals $p(x_j | y_i)$ are straightforward implementations of the probability tables we constructed from the Protea Atlas Project (section 3.2). The prior $p(y_i)$ is a distribution over species before any attributes are observed. We may view the output of our image classification network as such a prior, since it does not carry any information of the attributes. Of course, the output of the image classifier may in turn be viewed as the combination of a prior and evidence of image data.

The model in (2) can be altered easily to incorporate a subset of observed attributes (or none at all, in which case we simply return the prior $p(y_i)$).

5. Results and discussion

The aim of this work is to provide a dataset for *Protea* identification and to establish a baseline solution for this task. We focus specifically on gauging the individual and joint effects of the various components of our proposed model, which would hopefully be useful for future improvements or to solve similar problems. To this end, we proceed to report on the test performance of various versions of our model. All classification CNNs are trained with cross-entropy loss and the Adam optimiser with its default learning rate of 0.001. The λ and α parameters in equation (1) are set to 2 and 0.02, as recommended in [38]. No further hyper-parameter optimisation is performed for any of the networks.

Performance is measured in three ways: (1) top-1 accuracy, which is simply the ratio of correctly identified species over the entire test set; (2) top-3 accuracy, which is the ratio of test samples for which the correct species appeared in the model’s top three scores (useful in a semi-automated, recommender-type environment); and (3) recall, which in our context is average per-class accuracy. Recall ignores the class imbalance, and gives a better indication of whether rare species are correctly identified.

As a starting point we replaced and trained the fully-connected layers of a standard Inception-V3 network, as explained at the beginning of section 4.1. Here we implemented early stopping in two ways: one favouring high accuracy and one favouring high recall on the test set. Note that this is the only place where the test set is used for validation. We do so only to establish a baseline, not to use these models ever again. Results are shown in the first few lines of Table 1 (where a random classifier taking the class imbalance into account is also evaluated). Accuracy is almost double the recall, indicating that this network might have a bias for the more commonly occurring species.

The CNN-A model in Table 1 includes an attention mechanism, and performs markedly better than the previous model in terms of top-1 accuracy and recall (though not much in terms of top-3 accuracy, which is perhaps interesting).

The subgenus network described in section 4.2 on its own achieves a top-1 accuracy of 66.15% and a recall of 44.21%. These values are not directly comparable to those in Table 1, since the subgenus network solves a different problem. That said, it is an easier problem and we would expect performance to be relatively high.

The manner in which attributes are used in section 4.3 requires a prior. We experimented with a uniform prior (which gives a purely attribute-based classifier), and also priors obtained from the CNN-A network and the subgenus network (Subg). It might be worth noting that the pure attribute-based classifier performs similar to the standalone CNN without attention.

We also experimented with various combinations of the attributes, CNN-A model that classifies images on a species level, and the Subg network that classifies images on subgenus level. It is clear that the combination of all three components outperforms all other versions. We note that the inclusion of an attention mechanism in the species-level CNN, and also the incorporation of attributes, impact performance significantly. The effect of the subgenus-level CNN is slightly less, but still useful.

The per-class accuracies obtained with our full model on the test set can be seen in Figure 6. The class IDs in this graph are ordered the same as those in Figure 2, and it is encouraging to see that the model is able to identify rarer species with an accuracy more-or-less similar to that of more common species.

Figure 7 shows a few correct and incorrect identifications made by the full model on test images. The high degree of visual similarity between species like *Protea neriifolia* and *Protea lepidocarpodendron* is evident. The amount of training data for *Protea burchellii* is roughly six times less

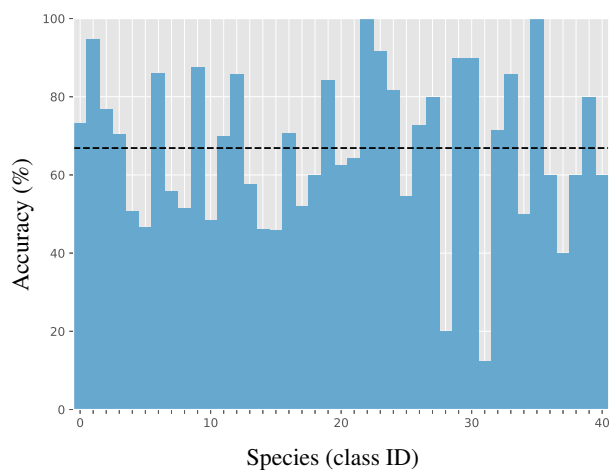


Figure 6. The per-class test accuracy from the final model shows no significant bias for the classes with more training images. The horizontal line indicates the average per-class accuracy (which is the recall value of 66.88% in Table 1).

Model		Top-1	Top-3	Recall
Random classifier		6.35%	18.11%	2.44%
CNN (without attention)		30.32%	59.29%	13.51%
CNN-A (with attention)		55.06%	77.15%	35.59%
Attr with uniform prior		25.86%	60.75%	34.84%
Attr with Subg prior		47.28%	76.78%	55.39%
Attr with CNN-A prior	<i>no Subg</i>	65.77%	83.35%	65.83%
CNN-A with Subg scaling	<i>no Attr</i>	56.73%	78.86%	35.43%
Attr with CNN and Subg prior	<i>no attention</i>	56.64%	80.45%	51.91%
Attr with CNN-A and Subg prior	<i>full model</i>	70.43%	85.80%	66.88%

Table 1. Test performance comparison of various versions of our model. The best performance is achieved by our full model that combines a CNN with attention, a subgenera network, and attributes.

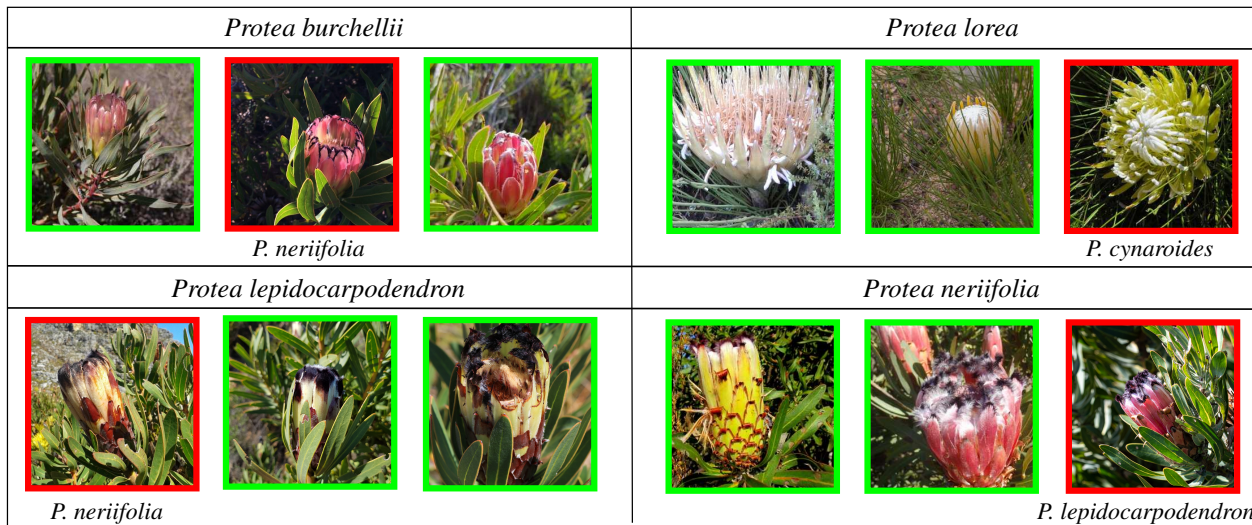


Figure 7. Example identification by our full model on four *Protea* species. The true label of each of the four species is shown above a set of test images, correctly identified images are outlined in green, and predicted labels are shown below incorrectly identified images.

than that of *Protea neriifolia*, so we may expect some of the latter to be confused for the former. *Protea lorea* has very few training images (only 22), yet is mostly correctly identified.

Figure 8 demonstrates the effects of the various attributes on the top 3 classification scores for a number of test images. *Protea cryophila* is localised to fewer than 10 high-altitude mountain peaks, and it is expected that the inclusion of location and elevation should impact its identification significantly. Similarly, *Protea effusa* is a high-altitude and highly localised species, and we observe a similar effect. For the example of *Protea lanceolata* the visual identification (without attributes) is already fairly certain, and the attributes further boosts the certainty in the desired way. The last example in the figure shows how the inclusion of all the attributes may still not be sufficient for the system to correctly identify the species. The inclusion of location does lead to the correct label appearing in the top 3 scores.

6. Conclusion

We considered the problem of *Protea* species identification from an image and optional information specifying the location, elevation and date of the observation (which we collectively refer to as attributes). The contribution of the paper is two-fold: we firstly introduce a challenging dataset for fine-grained image classification, and secondly propose an identification model that consists of a CNN with attention, a second CNN to classify on the coarser subgenera-level and rescale the output of the first CNN, and a probabilistic model to condition the identification on the observed attributes. The proposed combination of these three parts performs reasonably well on test data, and can form a basis for future studies.

As also noted in [31], datasets and studies like ours not only provide computer vision researchers with new challenges representative of the real world, but are also useful





	No attributes	Location	Elevation	Date	All attributes
<i>P. cryophila</i> 	<i>P. cynaroides</i> <i>P. longifolia</i> <i>P. nerifolia</i>	<i>P. cynaroides</i> <i>P. cryophila</i> <i>P. magnifica</i>	<i>P. cryophila</i> <i>P. magnifica</i> <i>P. punctata</i>	<i>P. cynaroides</i> <i>P. lorea</i> <i>P. magnifica</i>	<i>P. cryophila</i> <i>P. magnifica</i> <i>P. punctata</i>
<i>P. effusa</i> 	<i>P. nitida</i> <i>P. humiflora</i> <i>P. effusa</i>	<i>P. effusa</i> <i>P. amplexicaulis</i> <i>P. nitida</i>	<i>P. effusa</i> <i>P. sulphurea</i> <i>P. cynaroides</i>	<i>P. effusa</i> <i>P. nitida</i> <i>P. sulphurea</i>	<i>P. effusa</i> <i>P. amplexicaulis</i> <i>P. nitida</i>
<i>P. lanceolata</i> 	<i>P. lanceolata</i> <i>P. nitida</i> <i>P. repens</i>	<i>P. lanceolata</i> <i>P. repens</i> <i>P. nerifolia</i>	<i>P. lanceolata</i> <i>P. nitida</i> <i>P. repens</i>	<i>P. lanceolata</i> <i>P. nitida</i> <i>P. repens</i>	<i>P. lanceolata</i> <i>P. repens</i> <i>P. nerifolia</i>
<i>P. obtusifolia</i> 	<i>P. cynaroides</i> <i>P. repens</i> <i>P. lanceolata</i>	<i>P. susannae</i> <i>P. repens</i> <i>P. obtusifolia</i>	<i>P. cynaroides</i> <i>P. lanceolata</i> <i>P. susannae</i>	<i>P. cynaroides</i> <i>P. lanceolata</i> <i>P. repens</i>	<i>P. susannae</i> <i>P. obtusifolia</i> <i>P. compacta</i>

Figure 8. We compare the effects of including attributes on top 3 classification scores for four examples from the test set. “No attributes” means that only the CNN-A model on the image is used, and green indicates the correct label. The potentially important effect of the location and elevation attributes for *Protea* identification is apparent in the top two rows.

for a number of well-defined conservation and field biology purposes.

References

- [1] D. Apriyanti, A. Arymurthy, and L. Handoko. Identification of orchid species using content-based flower image retrieval. *International Conference on Computer, Control, Informatics and its Applications*, pages 53–57, 2013.
- [2] P. Barré, B. Stöver, K. Müller, and V. Steinhage. LeafNet: a computer vision system for automatic plant species identification. *Ecological Informatics*, 40:50–56, 2017.
- [3] S. Beery, G. van Horn, O. Mac Aodha, and P. Perona. The iWildCam 2018 challenge dataset. *arXiv preprint arXiv:1904.05986*, 2019.
- [4] T. Berg, J. Liu, S. Woo Lee, M. Alexander, D. Jacobs, and P. Belhumeur. BirdSnap: large-scale fine-grained visual categorization of birds. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2011–2018, 2014.
- [5] M. Chandler, L. See, C. Buesching, J. Cousins, C. Gillies, R. Kays, C. Newman, H. Pereira, and P. Tiago. Involving citizen scientists in biodiversity observation. In *The GEO Handbook on Biodiversity Observation Networks*, pages 211–237. Springer, 2017.
- [6] S. Cho. Content-based structural recognition for flower image classification. *IEEE Conference on Industrial Electronics and Applications*, pages 541–546, 2012.
- [7] S. Cho and P. Lim. A novel virus infection clustering for flower images identification. *International Conference on Pattern Recognition*, pages 1038–1041, 2006.
- [8] R. Cowling, R. Pressey, M. Rouget, and A. Lombard. A conservation plan for a global biodiversity hotspot — the Cape Floristic Region, South Africa. *Biological Conservation*, 112(1-2):191–216, 2003.
- [9] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: a large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.
- [10] J. Fu, H. Zheng, and T. Mei. Look closer to see better: recurrent attention convolutional neural network for fine-grained image recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4476–4484, 2017.
- [11] K. Gaston and M. O’Neill. Automated species identification: why not? *Philosophical Transactions of the Royal Society of*

- London. *Series B: Biological Sciences*, 359(1444):655–667, 2004.
- [12] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.
- [13] S. W. Hong and L. Choi. Automatic recognition of flowers through color and edge based contour detection. *International Conference on Image Processing Theory, Tools and Applications*, pages 141–146, 2012.
- [14] T. Hsu, C. Lee, and L. Chen. An interactive flower image recognition system. *Multimedia Tools and Applications*, 53(1):53–73, 2011.
- [15] R. Huang, S. Jin, J. Kim, and K. Hong. Flower image recognition using difference image entropy. *International Conference on Advances in Mobile Computing and Multimedia*, pages 618–621, 2009.
- [16] J. Krause, B. Sapp, A. Howard, H. Zhou, A. Toshev, T. Duerig, J. Philbin, and L. Fei-Fei. The unreasonable effectiveness of noisy data for fine-grained recognition. *European Conference on Computer Vision*, pages 301–320, 2016.
- [17] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.
- [18] T. Lin, A. Roychowdhury, and S. Maji. Bilinear cnn models for fine-grained visual recognition. *IEEE International Conference on Computer Vision*, pages 1449–1457, 2015.
- [19] M. Nilsback and A. Zisserman. A visual vocabulary for flower classification. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1447–1454, 2006.
- [20] M. Nilsback and A. Zisserman. Automated flower classification over a large number of classes. *Indian Conference on Computer Vision, Graphics and Image Processing*, pages 722–729, 2008.
- [21] M. Nilsback and A. Zisserman. Delving deeper into the whorl of flower segmentation. *Image and Vision Computing*, 28(6):1049–1062, 2010.
- [22] K. Phyu, A. Kutics, and A. Nakagawa. Self-adaptive feature extraction scheme for mobile image retrieval of flowers. *International Conference on Signal Image Technology and Internet Based Systems*, pages 366–373, 2012.
- [23] W. Qi, X. Liu, and J. Zhao. Flower classification based on local and spatial visual cues. *IEEE International Conference on Computer Science and Automation Engineering*, pages 670–674, 2012.
- [24] T. Rebelo. Protea Atlas. <http://hdl.handle.net/20.500.12143/5287>, 2004. Accessed: 2019-07-26.
- [25] T. Rebelo, C. Paterson-Jones, and N. Page. *SASOL Proteas: a field guide to the Proteas of Southern Africa*. Fernwood Press in association with the National Botanical Institute, 2001.
- [26] J. Silvertown. A new dawn for citizen science. *Trends in Ecology and Evolution*, 24(9):467–471, 2009.
- [27] M. Simon and E. Rodner. Neural activation constellations: unsupervised part model discovery with convolutional networks. *IEEE International Conference on Computer Vision*, pages 1143–1151, 2015.
- [28] M. Sun, Y. Yuan, F. Zhou, and E. Ding. Multi-attention multi-class constraint for fine-grained image recognition. *Lecture Notes in Computer Science*, 11220:834–850, 2018.
- [29] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the Inception architecture for computer vision. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2818–2826, 2016.
- [30] L. Valente, G. Reeves, J. Schnitzler, I. Mason, M. Fay, T. Rebelo, M. Chase, and T. Barraclough. Diversification of the African genus *Protea* (Proteaceae) in the Cape biodiversity hotspot and beyond: equal rates in different biomes. *Evolution: International Journal of Organic Evolution*, 64(3):745–760, 2010.
- [31] G. Van Horn, O. Mac Aodha, Y. Song, Y. Cui, C. Sun, A. Shepard, H. Adam, P. Perona, and S. Belongie. The iNaturalist species classification and detection dataset. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 8769–8778, 2018.
- [32] J. Wäldchen and P. Mäder. Plant species identification using computer vision techniques: a systematic literature review. In *Archives of Computational Methods in Engineering*, volume 25, pages 507–543. Springer, 2018.
- [33] J. Wäldchen, M. Rzanny, M. Seeland, and P. Mäder. Automated plant species identification — trends and future directions. *PLOS Computational Biology*, 14(4), 2018.
- [34] B. Weinstein. A computer vision for animal ecology. *Journal of Animal Ecology*, 87(3):533–545, 2018.
- [35] S. Wu, F. Bao, E. Xu, Y. Wang, Y. Chang, and Q. Xiang. A leaf recognition algorithm for plant classification using probabilistic neural network. *IEEE International Symposium on Signal Processing and Information Technology*, pages 11–16, 2007.
- [36] H. Zawbaa, M. Abbass, S. Basha, M. Hazman, and A. Hasenian. An automatic flower classification approach using machine learning algorithms. *International Conference on Advances in Computing, Communications and Informatics*, pages 895–901, 2014.
- [37] C. Zhang, P. Zhou, C. Li, and L. Liu. A convolutional neural network for leaves recognition using data augmentation. *IEEE International Conference on Computer and Information Technology*, pages 2143–2150, 2015.
- [38] H. Zheng, J. Fu, T. Mei, and J. Luo. Learning multi-attention convolutional neural network for fine-grained image recognition. In *IEEE International Conference on Computer Vision*, pages 5209–5217, 2017.