California Wildflower Field Guide & New Species Name Generator

Lorraine Lin
Stanford University
lhl3135@stanford.edu
https://github.com/lhl3135/CS230

Abstract

An image classifier powered by AlexNet produced an accuracy of 90%+ on California wildflowers. Additionally, a species name generator produced plausible scientific names from wildflower images using a modified image captioning algorithm. A new performance metric was defined for generating species name – recall of the species' genus (or first word).

1 Introduction

Only 1% of the estimated 2 billion species living on Earth have been classified, leaving a staggering 99% of new species to be discovered and named, including many plants [1]. Field Guides for birds, spiders, and flowers have remained continuously in demand. Biologist have speculated that the reason for the longstanding popularity in bird-watching and identifying plants is to exercise our innate instinct for hunting and gathering in a modern setting. The purpose of this project is to encourage an interest in nature and conservation by creating tools to: 1) accurately identify California wildflower species from images, and 2) generate plausible scientific (Latin) names for newly discovered species from wildflower images.

2 Related work

The first popular US field guide was published in 1895 entitled "How to Know the Wildflowers" [2]. Today, the Yosemite Conservancy sells popular one sheet plastic cards for identifying High Sierra wildflowers [3]. Both show the enduring fascination with the biodiversity of native flora. Convolutional Nets can greatly increase the speed and accuracy objects are identified [4, 5]. However, there is no app specifically for California Wildflowers. Such a tool, particularly if paired with the ability to create new species names, has the potential to crowd-source discovery of new plant species while monitoring environmental change with global warming.

Google's Show and Tell Model was an exciting breakthrough in generating plausible captions using Inception V3 and LSTMs [6]. It predicts sequences of words to form captions from images. More recent research by Xu et al. showed that it is possible to generate classical Chinese poetry from images [7]. Their work used an image-based encoder (BI-GRU) and memory-based decoder, combined with a keyword extractor, to generate plausible poems that met criteria for key concepts as elusive as "poeticness, fluency, coherence, meaning and consistency". Their performance metric, recall rate of these key concepts, was adopted for this project.

3 Dataset and Features

For wildflower identification, two image datasets were combined: 1) "102 Category Flower Dataset" published by the Oxford Visual Geometry Group for common flowers, and 2) 32 categories of common California wildflowers scraped off the web, then manually validated using the California Native Plant Society website [8, 9].

To generate new names, a third *wider & shallower* dataset was collected by scraping 6101 species of North American wildflowers from www.WildflowerSearch.org. It averaged two images per species to represents a wide range of species names but was not intended for classification. Nonetheless, there are many images for each genus (first word in scientific name).

All features were RGB images in jpg format scaled down and padded to the size required by each classic model. AlexNet and VGG-16 requires inputs of 227x227x3 and 224x224x3 images respectively, which were normalized with the ImageNet mean. The Inception V3 model requires inputs of 299x299x3 to create encoding vectors of size 2048 as inputs into the RNN with LSTM units for training & generating new species names.

4 Methods

For wildflower identification, two classic models were used: AlexNet and VGG-16 (see Table 1) [4, 5]. For scientific name generation, a modified version of Google's Show and Tell was used. The third model links the Inception V3 to an RNN with LSTM units [6]. All three models minimized a cross-entropy loss function:

$$J(\theta) = \sum_i y_i \, log(\hat{y}_i) \qquad [1]$$

AlexNet: An implementation of AlexNet by Kratzert [10] was adopted with the following changes for this project: replace last fully-connected layer with 134 classes of common and wildflower categories, Adam optimizer switched from GD, save and restart option from checkpoints added, evaluation of test set added in addition to train & dev sets, detailed output of predictions by class for error analysis added, and learning rate decay (not used). Pretrained ImageNet weights were loaded, then the last three fully-connected layers fine-tuned for the flower data.

VGG-16: An implementation of VGG-16 by Moindrot [11] was adopted with some changes: 134 flower classes, option to save checkpoints with detailed output for error analysis, and evaluation on test set in addition to training & dev sets. VGG-16 with pretrained ImageNet weights had its last fully-connected layer trained for 10 epochs, then all layers further finetuned at a lower learning rate for another 5 epochs.

Additional methods used for this project to reach convergence include: 1) incrementally loading new data, 2) trading off the number of images with cleaner data, 3) reshuffling more than once if optimization started off poorly (using the same hyperparameters), and 4) strategically tuning part of the model before finetuning all layers for VGG-16 (detailed above).

Table 1: Model Parameters & Hyperparameters for Wildflower Identification

Model	Batch Size	Optimizer	Input Size	α_0	α decay?	β_1	β_2	Dropout*
AlexNet	256	Adam	227	1e-4	No	0.9	0.999	0.5
VGG-16	32	Vanilla-GD	224	1e-3/1e-5	Yes	-	-	0.5

*keep probability

Show & Tell Model: For generating scientific names, a version of Google's Show & Tell Model was implemented with changes (see Table2). The original version used the Inception V3 model linked to an RNN with LSTM units to generate captions from images for the MS COCO competition [6]. Wang froze the Inception model weights, then trained the RNN for a select vocabulary to generate new captions [12].

A modified version of Wang's repository was adopted with code changes made for this project. Specifically, caption words were replaced by sequential letters for wildflower species name of each image. Additional modifications were made to read from files, rather than pulling images from urls. Predictions were limited to two-word representations, where the first is the Latin genus (or generic) name and the second the species' specific name. The padded length of predictions was increased from 25 to 100 to capture the maximum length of scientific names in the datasets. Again, pretrained

COCO weights for the Inception model were frozen during training of the vector-to-sequence RNN. Finally, a beam search algorithm generated *letter-by-letter* the new "binomial nomenclature" (i.e. two-word Latin names) from images. A range of beam sizes were tests.

Table 2: Model Parameters & Hyperparameters for Species Name Generation

Model	Batch Size	Optimizer	Input Size	α_0	α decay?	Dropout*	Padded Length	Beam Size
Show & Tell (Frozen Inception +Trained RNN)	32	SGD	299	2.0	Yes	0.7	100	1,3,5,10

*keep probability

5 Results/Metrics/Discussion

<u>Wildflower Identification</u>: Accuracy was deemed a sufficient performance metric for the importance of this problem, as it has been widely used in other studies on the Oxford Flower Dataset allowing for easy comparison.

$$Accuracy = \frac{\#\,Correct\ Predictions}{Total\ \#\,Predictions} \qquad [2]$$

AlexNet produced results approaching the state of the art for classification of the Oxford dataset (94% accuracy) [13]. AlexNet has a total of ~60M parameters to train and produced good results relatively quickly once a good learning rate (α) was found (see Figure 1a) [4]. No further tuning of β_1, β_2 or ε was needed for the Adam Optimizer (see Table 3 and Figures 2 & 3 for results). A few data augmentation techniques were tested such as mirroring, random cropping, and Fancy PCA Data Augmentation. However, as accuracy was already very good, any marginal improvement in performance was not considered a high priority versus testing the VGG-16 model and/or generating new species names.

VGG-16 has ~138M parameters [5], which took much longer to train without producing better results with the same, relatively small dataset or the need for significantly more compute time (see Figure 1b).

Table 3: Model Performance for Wildflower Identification

Model	Dev/Test Description	Dataset Size Train/Dev/Test	Performance Metric	Epochs	Performance Train	Dev	Test
AlexNet	WF Only	11463 / 408 / 408	Accuracy	25	0.95	0.92	0.91
AlexNet	All Categories	9428 / 1424 / 1430	Accuracy	50	1.00	0.88	0.90
VGG-16	All Categories	9428 / 1424 / 1430	Accuracy	10+5	0.75	0.62	0.61

Figure 1: a) AlexNet Accuracy Plot, b) VGG-16 Accuracy Plot, c) Show & Tell RNN Loss(Train)

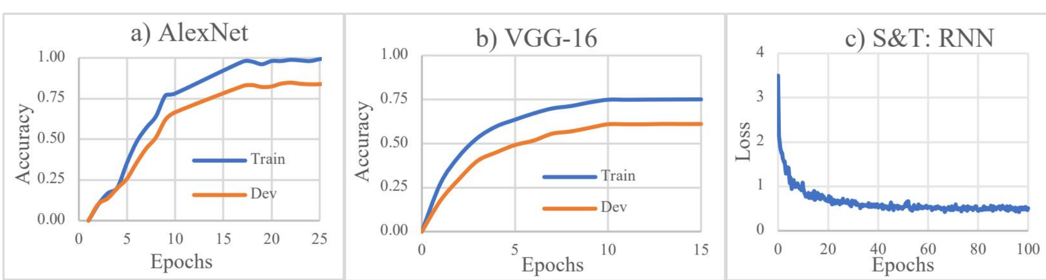


Figure 2: Error Analysis of Wildflower Identification

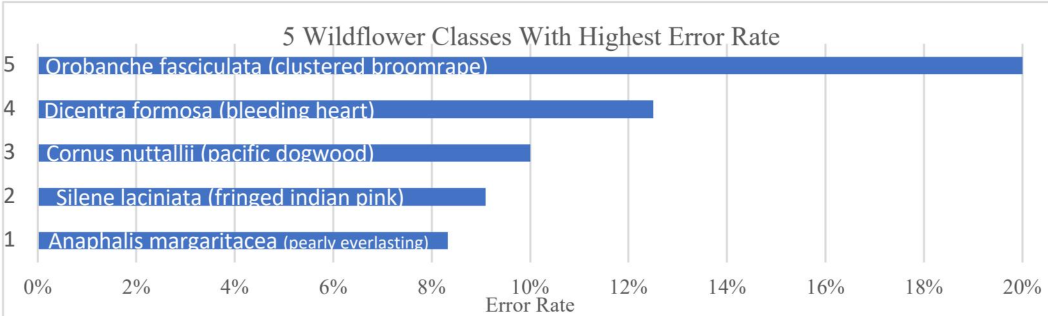


Figure 3: Two common misclassifications (from left to right): Silene laciniata (fringed indian pink) & Castilleja miniate (giant red paintbrush); Anaphalis margaritacea (pearly everlasting) & Cassiope mertensiana (white mountain heather)

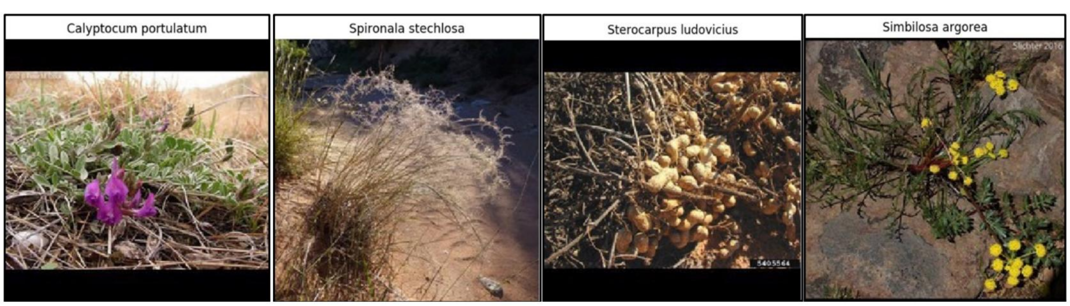
Generating Species Names: Unlike for machine translation or caption generation, the Bleu Score proved meaningless when calculated for a sequence of letters from generated and actual scientific names [14]. For new name generation, the dataset was selected to represent a wide range of species names, not to identify species. The RNN trained relatively quickly from the vector dataset (see Figure 1c). Beam sizes of 1, 3, 5 & 10 were tested, where a value equal to 1 (a greedy search) generated more unique names (see Figure 4). The model produced plausible Latin names, where the recall rate of the first word (or genus) was 16% for the dev & test sets (for beam size = 1), but less than 0.5% for the second (or specific name). Thus, the performance metric finally adopted was similar to that used for generating Chinese poetry from images: *recall rate of key concepts*, in this case recall of the *genus (or first word)* of the species.

$$Recall\ Rate = \frac{True\ Positives}{True\ Postives + False\ Negatives} \quad [3]$$

Table 4: Model Performance for Species Name Generation

Model	Dataset Size Train/Dev/Test	Beam Size	Performance Metric	Epochs	Performance Train	Dev	Test
Show & Tell Inception+RNN	19495 / 2436 / 2436	1	Recall of Genus	100	0.30	0.16	0.16

Figure 4: Algorithm output showing unique species names generated from their images



6 Conclusion/Future Work

- To generating more unique names, add another term to the loss function to penalize generating existing names.
- Incorporate more clean images using YOLO 9000 cropping. Oddly, many common flowers are not categories in the YOLO pretrained model [15].
- Create an app which has the wildflower identifier and species name generator linked into a continuous network.

7 Contributions

All ideas and code changes, with the exception of the references cited in the Methods Section, were made by the author.

References

[1] B Larsen, Brendan & Miller, Elizabeth & K Rhodes, Matthew & Wiens, John. (2017). Inordinate Fondness Multiplied and Redistributed: The Number of Species on Earth and the New Pie of Life. Quarterly Review of Biology. 92. 10.1086/693564.

[2] Parsons, F.T. (1895). How to Know the Wildflowers. Charles Scribner's Sons, New York.

[3] Morales-Denney, Elizabeth. (Date?). The Yosemite Conservancy, Sierra Nevada Field Card: Wildflowers.

[4] Krizhevsky, Alex & Sutskever, Ilya & E. Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

[5] Simonyan, Karen & Zisserman, Andrew. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv 1409.1556.

[6] Vinyals, Oriol & Toshev, Alexander & Bengio, Samy & Erhan, Dumitru. (2016). Show and Tell: Lessons learned from the 2015 MSCOCO Image Captioning Challenge. IEEE Transactions on Pattern Analysis and Machine Intelligence. 39. 1-1. 10.1109/TPAMI.2016.2587640.

[7] Xu, Linli & Jiang, Liang & Qin, Chuan & Wang, Zhe & Du, Dongfang. (2018). How Images Inspire Poems: Generating Classical Chinese Poetry from Images with Memory Networks.

[8] Nilsback, Aaria-Elena & Zisserman, Andrew. (2008) 102 Category Flower Dataset. http://www.robots.ox.ac.uk/~vgg/data/flowers/102/

[9] California Native Plant Society. https://www.cnps.org

[10] Kratzert, Frederik. (2016). Finetune AlexNet with Tensorflow, GitHub repository, https://github.com/kratzert/finetune_alexnet_with_tensorflow

[11] Moindrot, Olivier. (2017). Finetune VGG-16 with Tensorflow, GitHub repository, https://gist.github.com/omoindrot

[12] Wang, Weimin. (2017). Image Captioning Model in TensorFlow, GitHub repository, https://github.com/aaxwaz/Image-Captioning-Model-in-TensorFlow

[13] Do, Binh. (2018) Spatial Pyramid Pooling on Top of AlexNet Using Tensorflow, GitHub repository, https://github.com/peace195/sppnet

[14] Papineni, Kishore & Roukos, Salim & Ward, Todd & Zhu, Wei-jing. (2002). BLEU: A Method for Automatic Evaluation of Machine Translation. 10.3115/1073083.1073135.

[15] Redmon, Joseph & Farhadi, Ali. (2016). YOLO9000: Better, Faster, Stronger.