Using Active Learning to Reduce Training Data in Deep-learning-based Flower Classification

Feng Chen
School of Computer Science
Jubilee Campus
University of Nottingham
NG8 1BB
feng.chen@nottingham.ac.uk

Tony Pridmore
School of Computer Science
Jubilee Campus
University of Nottingham
NG8 1BB
tony.pridmore@nottingham.ac.uk

Andrew French
School of Computer Science
Jubilee Campus
University of Nottingham
NG8 1BB
andrew.p.french@nottingham.ac.uk

Conducting plant-related research using deep learning has become more and more popular in recent years. Deep learning is proving a powerful technique in many areas such as segmentation, counting and classification. However, to obtain a good deep learning model often requires much training data; while acquiring and annotating such data is expensive and time consuming. Crowdsourcing is a cheaper and faster way to get annotations [1], but these annotations can be noisy and unreliable because the expertise of the annotators is not guaranteed. Generative Adversarial Networks (GANs) [2] are also used for generating synthetic training data [3], but successfully training a GAN itself needs a lot of data. But maybe we are not using training datasets as efficiently as we might. Tuning the training approach using active learning is a potentially effective method to reduce the costs of training deep models.

Active learning can be used to select the most representative samples and to query information sources (e.g. humans and computers) for their annotations [4].A typical active learning dataset is formed by a small initial training set D_{train}, and a relatively large pooling set D_{pool} at the beginning, as iteratively more samples (i.e. pooled samples) are selected from the pooling set and added to the training set using the suitable acquisition functions. A pooled sample can be indicated as:

$$x^p = argmax_{x \in D_{pool}} a(x, M) \quad (1)$$

where x^p is the pooled sample, and $a(x, M)$, the acquisition function, is a function of x and the learning model M.

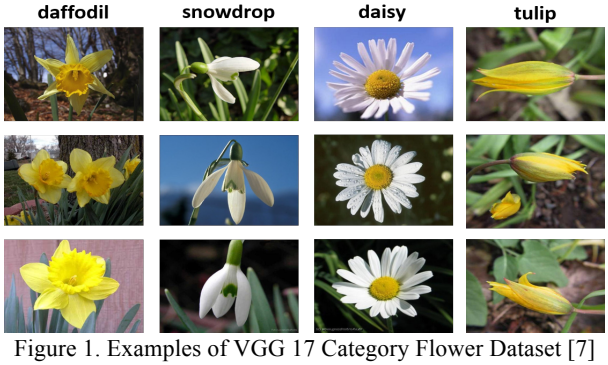

Figure 1. Examples of VGG 17 Category Flower Dataset [7]

In this paper we explore several different acquisition functions and evaluate them on a flower image dataset. Most of the functions try to acquire the most informative samples, including:

1. Random Selection (RS)

RS randomly selects new samples from the pool in each acquisition iteration. This method is used as a baseline in our experiments.

2. Max Entropy (Max E)

Max Entropy [5] selects the samples which *maximize their predictive entropy*. The predictive entropy is defined as:

$$H(y|x, M) = -\sum_c p(y = c|x, M) \log p(y = c|x, M) \quad (2)$$

where $x \in D_{pool}$, c is a given class, and $p(y|x, M)$ is the probability that x belongs to class y predicted by model M.

3. Max Variation Ratio (Max VR)

Variation ratio [6] is defined as:

$$VR(x) = 1 - \max_y p(y|x, M) \quad (3)$$

where $x \in D_{pool}$. Instead of using predictive entropy, the variation ratio takes the *highest predictive probability* of each sample to measure the uncertainty.

4. Min Standard Deviation (Min SD)

Min SD selects the samples which *minimize the standard deviation* across their predictive probabilities. This standard deviation is defined as:

$$SD(x) = \sqrt{\frac{1}{C}\sum_c [p(y = c|x, M) - \hat{p}(y|x, M)]^2} \quad (4)$$

where $x \in D_{pool}$, and $\hat{p}(y|x, M)$ is the average predictive probability of sample x.

A 17-category classification task is performed in our experiments using VGG 17 Category Flower Dataset (examples are shown in Figure 1) [7], which contains 80 images in each class (1360 images in total). We divide the dataset into 4 parts, including 68 (5%) initial training images, 884 (65%) pool images, 204 (15%) validation images, and 204 (15%) test images, and all of them are class-balanced.

In the pipeline shown in Figure 2, the deep learning models used in our experiments are modified VGG-16 and

VGG-19 [8]. The original output layer and the last fully connected layer are replaced by a logistic regression layer in each model. The ImageNet [9] pretrained weights are adopted, so in each model only the added logistic regression layer is trained. In each acquisition round, 17 new images are moved to the training set from the pooling set, and the logistic regression layer is trained on the newly formed training samples from scratch.

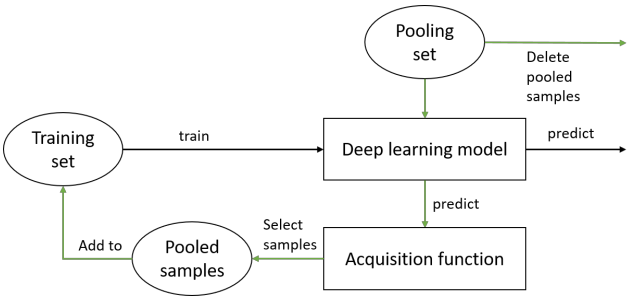

Figure 2. Active learning pipeline: the black arrows and the green arrows respectively denote the paths of the training samples and pooling samples.

Figure 3 shows our results after 35 acquisitions using different acquisition functions, compared to a baseline model without adding any acquisition function. The baseline model is trained on 85% of the data and tested on the same test set (15%) as used in other models. Our results indicate that when using proper acquisition functions, the number of training images is reduced by 43% (i.e. from 1156 to 663 images), while less than 2% of accuracy is decreased. Among the acquisition functions, random selection is the worst as expected, and Max E, Max VR and Min SD perform similarly.

Our best results from both models, Min SD in VGG-16 and Max VR in VGG-19, are compared to [10], where the same models (VGG-16/19 + logistic regression) are applied to the VGG 17 Category Flower Dataset without acquisition functions. The results, as presented in Table 1, again show the effectiveness of acquisition functions.

In conclusion, our research finds that applying active learning techniques can significantly reduce the number of training data in deep-learning-based flower classification tasks with little cost in accuracy. There is no reason to suppose this cannot be further expanded to other phenotyping tasks like leaf counting and segmentation when they are treated as classification. Given the high price and limited availability of good quality training sets, active learning should be an approach the phenotyping community is aware of.

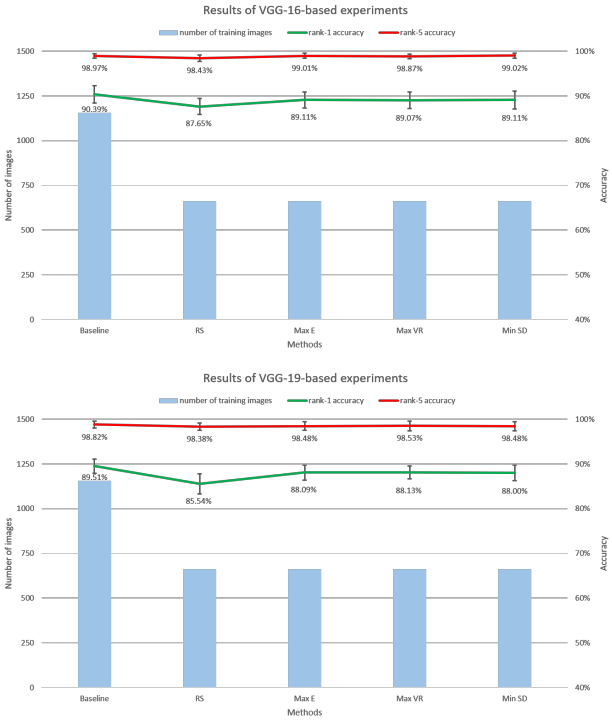

Figure 3. Results of VGG-16/19 with different acquisition functions: the graphs contrast the training size and performance between the baseline (i.e. without adding any acquisition function) and models with acquisition functions. Our results are reported in mean and standard deviation of rank-1 accuracy and rank-5 accuracy after 10 repeated experiments. In each graph, the rank-1/rank-5 percentage accuracies are reported in green/red lines with error bars (i.e. standard deviation) along the vertical axis on the right, and the numbers of training images are reported in blue bars along the vertical axis on the left.

Base model	Experiment	Training size	Test size	Rank-1 accuracy	Rank-5 accuracy
VGG-16	[10]	1224	136	85.29%	98.53%
		952	408	88.24%	99.02%
	Min SD	663	204	89.11%	99.02%
VGG-19	[10]	1224	136	88.24%	99.26%
		952	408	88.73%	98.77%
	Max VR	663	204	88.13%	98.53%

Table 1. Best results of our models (in blue background) compared to the ones in [10]: to be noticed, the training size and test size in [10] are different from ours.

References

[1] M. V. Giuffrida, F. Chen, H. Scharr, and S. A. Tsaftaris. Citizen crowds and experts: observer variability in image-based plant phenotyping. *Plant Methods*, vol. 14, p. 12, 2018.

[2] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. *In Advances in Neural*

Information Processing Systems (NIPS) 2014, pages 2672-2680, 2014.

[3] Y. Zhu. Data Augmentation using Conditional Generative Adversarial Networks for Leaf Counting in Arabidopsis Plants. In *CVPPP - BMVC 2018 Workshop*, 2018.

[4] S. Das, W. Wong, T. Dietterich, A. Fern, and A. Emmott. Incorporating Expert Feedback into Active Anomaly Discovery. In *IEEE 16th International Conference on Data Mining (ICDM),* 2016.

[5] C. Shannon and W. Weaver. The Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

[6] L. C. Freeman, *Elementary Applied Statistics: For Students in Behavioral Science*. Wiley, 1965.

[7] M. E. Nilsback and A. Zisserman. A Visual Vocabulary for Flower Classification. In *CVPR,* 2006.

[8] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *International Conference on Learning Representations (ICLR),* 2015.

[9] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei. ImageNet: A large-scale hierarchical image database. In *CVPR,* 2009.

[10] I. Gogul and V. S. Kumar. Flower species recognition system using convolution neural networks and transfer learning. In *Fourth International Conference on Signal Processing, Communication and Networking (ICSCN)*, 2017.