

Project Short Proposal

Luke Trinity, Nat Shenton, & Sean Klink
CS 254

September 26, 2019

1 Introduction

Online sources of high resolution images including Google, Image-net [1], and Flickr are providing new opportunities to explore applications of image categorization using computer vision. One interesting application in the field of fine-grained image classification is identification of dog breed. Dogs faces are considered to be highly differentiable for each breed, which makes classification an approachable problem. However, the work is challenging due to the varying combinations of facial characteristics and color patterns between species, as well as intra-class variation within a single species. Another aspect of the problem that makes it more complicated is the presence of humans and man-made backgrounds, which are not as common in other animal datasets. In addition, many attempts to classify dog breeds fail to corroborate the results with actual DNA evidence. Our approach will be to utilize an existing dog breed image classification dataset [2], in combination with a novel dataset that includes paired DNA test results and photos [3], to classify non-pure bred dog breeds.

2 Problem Definition and Algorithm

First we will convert each of the images in the Stanford Dogs dataset [2] into a single feature vector based on the trimmed height and width of the images and the (r,g,b) values in each cell. We will then pair the images with their associated class labels, and utilize a neural network as a black box to train our model to accurately predict the class of our test images. Later in the semester we hope to unbox the neural network with more advanced knowledge to achieve better accuracy. One method we will work to implement, which was utilized by both Kohsla et al. [2] and Liu et al. [4], is scale-invariant feature transform (SIFT). This algorithm detects and describes local features in images by finding candidate matching features based on Euclidean distance of feature vectors. [5] After achieving a high level of accuracy classifying pure-bred dogs, we will introduce our other dataset [3] that carries interesting practical concerns in our implementation of SIFT. Here we will attempt to classify mixed breed dogs using images annotated with precise DNA testing of breed composition. Hopefully we can augment this dataset with more samples by reaching out to different labs. This is an interesting computer vision problem because dogs often share features, whether it be color patterns or facial structure, that lead to breed misrepresentation. Training an algorithm in this capacity could lead to novel conclusions related to classifying a dog as purebred or a mixed breed.

3 Dataset

The Stanford Dogs Dataset has 20,580 images each classified into one of 120 classes of dog breed. [2] Each image also has an associated bounding box identifying the location of the dog within the photo. The mixed-breed DNA identified dog photo dataset from Voith et al. [3] has 20 images, each paired with DNA percentage breakdown of dog breed. There are no associated bounding boxes for the Voith dataset [3]. We will need to outfit this second dataset with bounding boxes to improve the accuracy of our classification algorithm, when the second phase of the project is reached. Later in the semester we hope to have other mixed breed training data, our backup plan is to scrape dog show images hosted online to generate our own dataset.

4 Related Work

In the initial article utilizing the novel dataset by Khosla et al. [2], the mean accuracy reached 22% when 100 training images were utilized within the SIFT methodology. Liu et al. [4] reached 69% accuracy, with greater success than Khosla et al. [2] due to their partially localized approach. We hope to leverage the bounding boxes and augment the work of Liu et al. [4], possibly improving on their methodology to achieve higher accuracy. Several bounding box approaches are currently being reviewed, although all are endlessly modifiable from simple image segmentation. One is a slight adaptation to a method often employed for training algorithms to images of different shade and dimension called DeepCut. [6] There are several standard variations on the basic DeepCut methodology, which hones in on the nature and dimension of the bounding in a cookie-cutter, traced-out way as opposed to any kind of geometric shape. A very similar methodology is known as GrabCut which does not need to even be a single, continuous cookie-cutter-style line bounded around a desired segment of an image. GrabCut is standardly designed to grab multiple disparate segments of an image at once for a single, desired analysis. In addition to one of these approaches or some modification thereof, we plan to incorporate some aspect of KL Loss to account for localization uncertainty which could greatly assist with interference from background objects such as people, buildings, trees, etc. [7] There are, of course, a number of other modifiable ‘standard’ or ‘familiar’ combinations of approaches to deal with bounding and uncertainty to also be explored. Even so, most will lead to an approximately similar result if implemented correctly. Deep/GrabCut is simple and can narrow the selected image segment in a convenient, efficient manner. KL Loss will be further researched and incorporated to deal with expected shading difficulties, image blurring and background/foreground image ‘noise’.

References

- [1] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009.
- [2] A. Khosla, N. Jayadevaprakash, B. Yao, and F.-F. Li, “Novel dataset for fine-grained image categorization: Stanford dogs,” in *Proc. CVPR Workshop on Fine-Grained Visual Categorization (FGVC)*, vol. 2, 2011.
- [3] V. L. Voith, E. Ingram, K. Mitsouras, and K. Irizarry, “Comparison of adoption agency breed identification and dna breed identification of dogs,” *Journal of Applied Animal Welfare Science*, vol. 12, no. 3, pp. 253–262, 2009.
- [4] J. Liu, A. Kanazawa, D. Jacobs, and P. Belhumeur, “Dog breed classification using part localization,” in *European conference on computer vision*, pp. 172–185, Springer, 2012.
- [5] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [6] M. Rajchl, M. C. Lee, O. Oktay, K. Kamnitsas, J. Passerat-Palmbach, W. Bai, M. Damodaram, M. A. Rutherford, J. V. Hajnal, B. Kainz, *et al.*, “Deepcut: Object segmentation from bounding box annotations using convolutional neural networks,” *IEEE transactions on medical imaging*, vol. 36, no. 2, pp. 674–683, 2016.
- [7] Y. He, C. Zhu, J. Wang, M. Savvides, and X. Zhang, “Bounding box regression with uncertainty for accurate object detection,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2888–2897, 2019.