CS230

Dog Breed Classification and Visualization

Jesse Candido
Department of Computer Science
Stanford University
jcandido@stanford.edu

Abstract

When someone gets their dog from shelters or from the streets, it can be difficult to know their pet's breed. This piece of information can be interesting to know, and even important. This project uses a CNN to identify the breed of a dog from a photo. The test accuracy was 84.21%.

1 Introduction

Approximately 6.5 million pets enter US animal shelters every year, according to the American Society for the Prevention of Cruelty to Animals. We domesticate a species, taking away their freedom and independence, and many of them don't even get the one thing they should be getting – companionship. Of the pets entering shelters each year, 3.3 million are dogs. Some are adopted, some are euthanized, some are abandoned to suffering. When approximately 44% of all households in the US have a dog [ASPCA], only 23% of people get their dog from a shelter, and 6% are stray. Most dogs (46%) are coming from breeders and private parties. When people start paying for dogs, they are incentivizing an economy that protects a few breeds while leaving many others without a home.

When someone decides to adopt a dog from a shelter or the streets, however, it is not uncommon to be unsure about its breed. In many cases it is impossible to ask someone about its pedigree, unless the dog comes from one of the few popular and distinctive breeds such as a golden retriever. It can be really important to know the breed of the dog. Some breeds, for example, evolved to perform certain tasks, and this knowledge would be beneficial for training (for example, hunting retrievers search for items and either bring them or hide them away). The pet could have some health issues that the owner and veterinarian should be aware of, that are associated with its breed (Great Danes, for example, are prone to Addison's disease [1]). In addition, understanding the breed can help the owners stay safe (some breeds were bred to hunt and kill and are more likely to bite, for example).

This project is an application of convolutional neural networks. It can be quite challenging to identify a breed. Some dog breeds are similar to each other that even humans have a hard time telling them apart. Some breeds have dogs with different colors, and in some breeds they all look the same. A good algorithm, therefore, has to be able to identify both a gold Labrador and a black Labrador as part of the same class. There is a great variation (in color, size, ear position, fur length, pose etc.) across breeds, but also within breeds, which makes this task challenging. My goal for this project was to achieve accuracy of around 80%.

The input to the algorithm is an image of a dog. We then use a convolutional neural network to output a predicted breed.

CS230: Deep Learning, Spring 2018, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

2 Related work

Liu J., Kanawaza A., Jacobs D., and Belhumeur P. (2012) in their paper "Dog Breed Classification Using Part Localization" achieved a 67% recognition rate by identifying face parts (eg. face and eyes). To do so, they had to build geometric models of dog breeds. Prasong, Pusit, and Kosin Chamnongthai, in their paper "Face-Recognition-Based dog-Breed classification using size and position of each local part, and pca," also identify sizes and positions of facial features to perform classification. Similarly, in the paper "Cats and Dogs," by Parkhi, Omkar M., et al. (2012), the authors identify breeds of cats and dogs based on the shape and appearance of the animal. While the results were successful, the categorization involved only 37 different breeds. This approach of identifying parts and distances between facial features is smart and accurate, but requires a complex categorization of features of different breeds.

The paper "Bird Species Categorization Using Pose Normalized Deep Convolutional Nets" by Van Horn, G., Branson S., Belongie S., and Perona P. (2014), on the other hand, uses a convolutional network to classify bird species, but before inputting the images to the network, they estimate the object's pose, and the image features are then used for classification. A CNN trained on many images can learn to "identify" edges and recognize objects while ignoring the background, which allows us to have a minimal alteration of the input. David Hsu, in the paper "Using Convolutional Neural Networks to Classify Dog Breeds," uses LeNet and GoogLeNet to perform classification. Yijia Hao, on the other hand, also performs transfer learning to identify dog breeds, but using ResNet50 instead of Xception.

3 Dataset and Features

The dataset used was of 133 dog breeds from the Stanford Dogs Dataset [2] and another dataset with American Kennel Club (AKC) recognized dog breeds. The images in these datasets were downloaded from Image-net, Google and Flickr. In total, there were 8351 original images, all in color and with the correct breeds as their label. This is what typical images in the dataset look like:

Figure 1: Image of German Pinscher

For data augmentation, images within a breed were randomly selected and then augmented with three different approaches:

- Flipping the image horizontally
- Adding random rotation to the left or right
- Adding random noise

The augmented dataset had a total of 21651 images (Total training images: 19980. Total validation images: 835. Test images: 836).

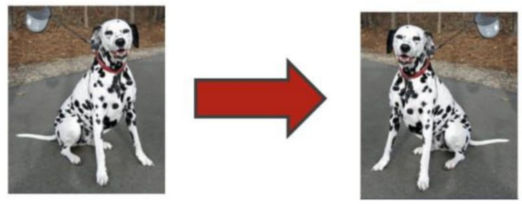

Figure 2: Image of Dalmatian flipped horizontally

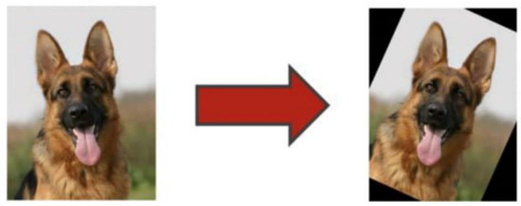

Figure 3: Image of German Sheppard with random rotation

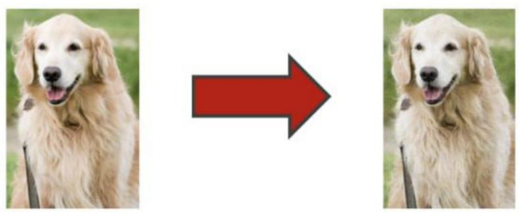

Figure 4: Image of Golden Retriever with random noise

The data was preprocessed to input a 4D array. The images were resized to a square image that is 224x224 pixels, and then the image is converted to an array. The images are rescaled by dividing each pixel by 255.

4 Methods

- Simple (conv2d-batch-relu-maxpool-dropout)

First, a Convolutional Neural Network was trained from scratch, using Keras: conv2d-batch-relu-maxpool-dropout (repeated 4 times), adding dropout at the end to reduce overfitting. This led to a test accuracy of 11%.

- Xception

With transfer learning (using the pre-trained model Xception trained on ImageNet), and a final softmax layer, a much better performance was achieved.

The loss function used was cross entropy. The formula for cross entropy is:

5 Experiments/Results/Discussion

$$H_{y'}(y) := -\sum_i y'_i \log(y_i)$$

Figure 5: Formula for cross entropy, where y'_i is the true label for the ith training instance, and y_i is the predicted result of the classifier for the ith instance. This loss is what we want to minimize during training.

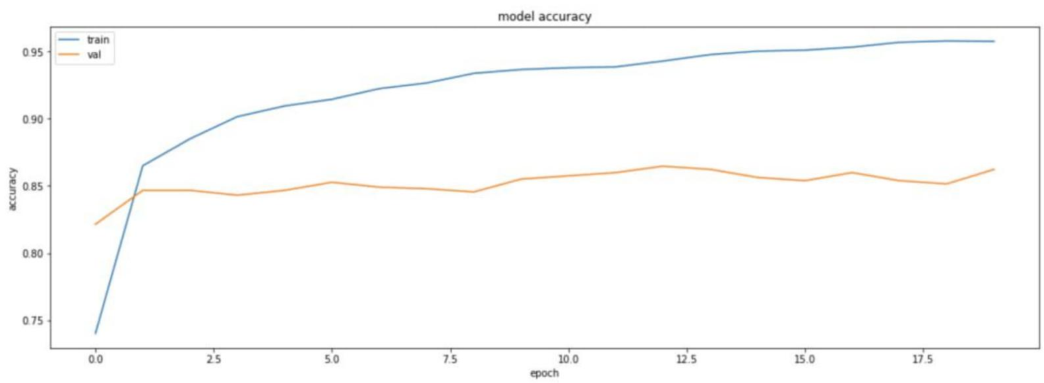

Figure 6: Plot of the model accuracy (using Xception and a final softmax layer) in the training and validation sets. The test accuracy for the best weights was 84.21%.

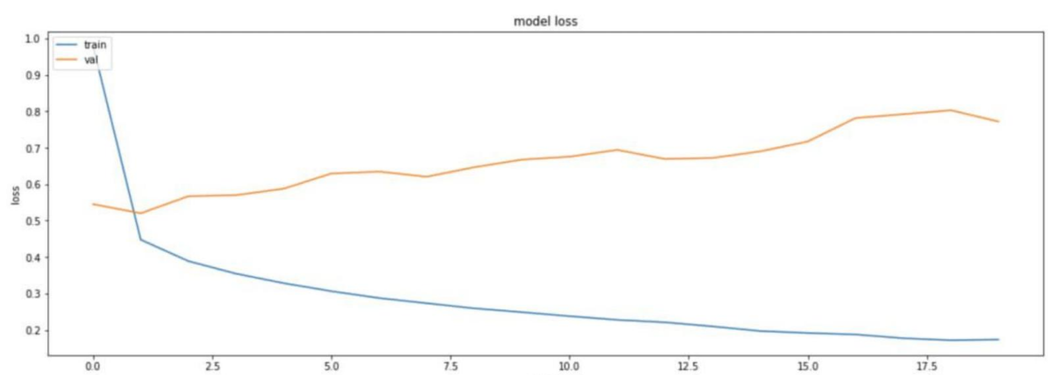

Figure 7: Plot of the loss for the training and validation sets.

Model	Time for one epoch (s)	Train Acc (%)	Val Acc (%)	Test Acc (%)
Simple	170	11.3	8.5	11.04
Xception	8	85.3	82.5	84.21

Figure 8: Table comparing the models tested.

There was a large difference between training and test/val accuracy. This is likely because the size of the dataset is still not large enough. Regardless, an accuracy of over 80% is great considering that we have 133 breeds and many are quite similar even to the human eye. The batch size used was 32, for training stability so that the learning process would converge quickly. The primary metric was accuracy.

To understand what the network is “seeing,” and to make sure that we are identifying the dogs and not the environment (in my subjective perspective, some breeds seemed to have more photos taken in a grass field, for example), we can use partial occlusion, in which the darker areas are the most important:

Figure 9: Figure showing the results of partial occlusion on a photo. The background regions are very bright because occluding the background doesn’t affect the probability of this dog being classified to the correct breed.

6 Conclusion/Future Work

Using a model pre-trained on ImageNet, the accuracy was significantly better than the simple model. The Xception model "maps the spatial correlations for each output channel separately, and then performs a 1x1 depthwise convolution to capture cross-channel correlation." [8]

In the future, further augmenting the dataset would be ideal. Understanding the physical appearance of dogs of mixed breeds would be interesting. If you mix two given breeds, would the neural network identify the dog as the original breeds with high probability? In addition, it could be worthwhile to investigate breeds that look too alike for humans and even for the neural network. What traits are similar? How do they differ?

References

[1] Waggener, Natalie, "Five Reasons Why Knowing Your Dog’s Breed Is Important," https://www.southbostonanimalhospital.com/blog/five-reasons-why-knowing-your-dogs-breed-is-

important

[2] Aditya Khosla, Nityananda Jayadevaprakash, Bangpeng Yao and Li Fei-Fei. Novel dataset for Fine-Grained Image Categorization. First Workshop on Fine-Grained Visual Categorization (FGVC), IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011

[3] ImageNet, “ImageNet.” www.image-net.org/. Accessed 10 June 2018.

[4] Yijia Hao, "Dog Breed Classification" (2017).

[5] A. Krizhevsky, I. Sutskever, and G. Hinton. “Imagenet classification with deep convolutional neural networks”, Advances in neural information processing systems. 2012.

[6] arXiv:1610.02357 [cs.CV]

[7] Henderson, R. & Rothe, R., (2017),"Picasso: A Modular Framework for Visualizing the Learning Process of Neural Network Image Classifiers." Journal of Open Research Software. 5(1), p.22.

[8] Xu, Joyce, "An Intuitive Guide to Deep Network Architectures," https://towardsdatascience.com/an-intuitive-guide-to-deep-network-architectures-65fdc477db41

[9] F. Chollet, “Keras,” https://github.com/fchollet/keras.

[10] Liu, Jiongxin, et al. "Dog breed classification using part localization." European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2012.

[11] Van Horn, G., Branson S., Belongie S., and Perona P. (2014) Bird Species Categorization Using Pose Normalized Deep Convolutional Nets. arXiv:1406.2952 [cs.CV]

[12] Parkhi, Omkar M., et al. "Cats and dogs." Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012.

[13] Prasong, Pusit, and Kosin Chamnongthai. "Face-Recognition-Based dog-Breed classification using size and position of each local part, and pca." Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2012 9th International Conference on. IEEE, 2012.

[14] Hsu, David. "Using Convolutional Neural Networks to Classify Dog Breeds," https://pdfs.semanticscholar.org/d58a/663d775c3ef6398d605b6d6dbd02fb3d8725.pdf.