Dog Breed Classification and Visualization

Jesse Candido jcandido@stanford.edu

Stanford Computer Science

Introduction

Approximately 3.3 million dogs enter US animal shelters every year. While 44% of all households in the US have a dog, only 23% of people get their dog from a shelter. When someone adopts a dog from a shelter or the streets, however, it is not uncommon to be unsure about its breed. This knowledge could be useful for the health of the pet, its training and even the safety of the owner. With this issue in mind, I built a Convolutional Neural Network to identify a breed by looking at a photo of the dog, achieving a test accuracy of 84.21%.

Data

The dataset used was of 133 dog breeds from the Stanford Dogs Dataset 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. For data augmentation, images within a breed were randomly selected and then augmented with three different approaches:

- Flipping the image horizontally

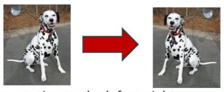

Image of Dalmatian flipped horizontally

- Adding random rotation to the left or right

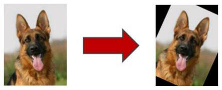

Image of German Sheppard with random rotation

- Adding random noise

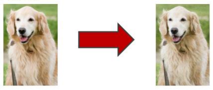

Image of Golden Retriever with random noise

The augmented dataset had a total of 21651 images.

Features

The data was pre-processed 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. There was a total of 272517 parameters in the final model.

```
Layer (type)                 Output Shape              Param #
=================================================================
global_average_pooling2d_1 ( (None, 2048)              0
_________________________________________________________________
dense_1 (Dense)              (None, 133)               272517
=================================================================
Total params: 272,517
Trainable params: 272,517
Non-trainable params: 0
_________________________________________________________________
```

Models

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

First, a Convolutional Neural Network was trained from scratch, using Keras: sequential model, conv2d-batch-relu-maxpool-dropout, 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. Xception consists of 1x1 convolution filters, followed by multiple 3x3 or 5x5 filters.

Results

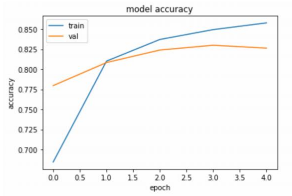

Figure 1 demonstrates the model accuracy (using Xception and a final softmax layer) in the training and validation sets. After a few epochs, the accuracies stay stable. The test accuracy for the best weights was 84.21%.

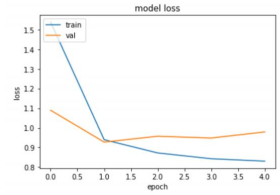

Figure 2 demonstrates the loss for the training and validation sets.

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

Total training images: 19980. Total validation images: 835. Test images: 836

Discussion

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. 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:

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]https://www.southbostonanimalhospital.com/blog/five-reasons-why-knowing-your-dogs-breed-is-important
[2] http://vision.stanford.edu/aditya86/ImageNetDogs/
[3]http://legacydirs.umiacs.umd.edu/~kanazawa/papers/eccv2012_dog_final.pdf
[4] https://www.kaggle.com/keras/resnet50
[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.