Dog Breed Classification

CS 5891: Final Project

Alex Feeley
Department of Electrical Engineering and Computer Science
Vanderbilt University

***Abstract*—Many animal shelters attempt to determine a dog's genetic make-up using only their visual appearance. It's hypothesized that a a deep learning model may be a better predictor of a dog's breed than shelter staff, which may aid in adoption efforts. A two-layer convolutional neural network was used as a baseline for this project. A pre-trained ResNet50 model was then used to achieve better results. Overall, the ResNet50 model achieved an accuracy of over 91% for classifying images of all 133 dog breeds.**

I. Introduction

Most animal shelters use the visual appearance of dogs to determine their breed. As most shelter dogs are mixed-breed, this becomes even more complicated. In a study at the Arizona Welfare League & SPCA (AAWL) in Phoenix, Arizona, it was found that shelter staff could only identify more than one of the breeds contained in a dogs' genetic make-up 10.4% of the time [2]. When shelter staff are able to correctly identify a dog's full genetic make-up, they will be more aware of a dog's physical, behavioral, and health characteristics that are associated with its breed. This can greatly aid in adoption efforts.

Most of the time shelter staff rely on a dog's size, weight, musculature, legs, coat, and tail to determine its breed. It is hypothesized that a deep learning model should be able to better classify dogs into their respective breeds based on their appearance. Additionally, a deep learning model may be able to better idenitfy a number of breeds that are part of the genertic make-up for a specific dog. This can allow staff to focus their adoption efforts on dogs with Pitbull ancestry, as Pitbul-type dogs are likely to have an average shelter stay of 37.5 days in comparison to non Pitbul-type dog's average length of stay of 19.7 days [2].

Overall, the identification of the genetic make-up of a dog can help in its adoption efforts. This issue is important to me because I volunteer with, and foster dogs on a regular basis. I think it'd help potential adopters to know the behavioral and health characteristics which surround a specific breed.

The goal of this project was to develop a dog breed classification system which could identify images of dogs as one of 133 breeds. This system can also give the top-5 breeds which a specific dog image may be classified into, which can help in identifying images of mixed-breed shelter dogs.

In the future, I hope to improve this model and make it into a simple application to aid shelter staff and adopters in their adoption efforts of shelter dogs.

II. Data Pre-Processing

The data is sourced from the Udacity dog project [3]. In total, there were 8,351 images which were divided among 133 dog breeds. The data was used in an 80/10/10 split for training, validation, and testing data, resulting in 6,680 training images, 835 validation images, and 836 testing images.

All images were pre-processed before use. Images were resized to 256 by 256 tensors. A center crop was then used to get a final image tensor size of 224 by 224. These tensors were then normalized using mean $= [0.485, 0.456, 0.406]$ and std $= [0.229, 0.224, 0.225]$ according to ImageNet standards [1]. Additionally, data augmentaion was used for the training data. The training data was randomly cropped to 256 by 256 instead of resized, randomly rotated by 15 degrees, used color jitter, and randomly had horizontal flips. Overall, this resulted in a more robust training set that was used to train the model. The data augmentation of the image of a beagle is shown in Figure 1.

III. Baseline Model

A. Architecture

The baseline model for this project was a two-layer convolutional neural network. Each convo-

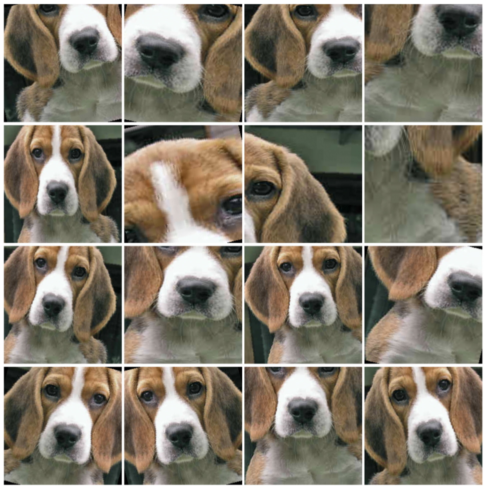

Fig. 1. A single picture of a Beagle was augmented in 16 separate images for the training data set

lutional layer used the ReLU activation function and was followed by a max-pooling layer. The convolutional layers used a kernel size of 5 and a stride of 1. The number of input and output channels can be seen in Figure 2. These values were chosen after much success in the loss and accuracy from homework 4. The max-pooling layers used a kernel size of 2 and a stride of 2. The two convolutional layers were followed by three fully connected layers, which resulted in 133 output neurons. This allowed the images to be classified into one of 133 categories, or breeds, using one-hot encoding.

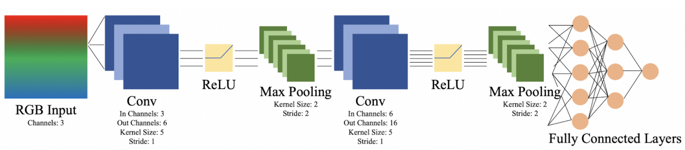

Fig. 2. Two-Layer Convolutional Neural Network used as a Baseline Model

B. Results

The baseline model achieved a validation accuracy of almost 60%. Due to data augmentation, the training accuracy was only 25%. However, as can be seen in Figure 3 and Figure 4, these results were only achieved after almost 450 epochs. The accuracy for the test set was 57%, which is near identical to the accuracy on the validation set. Overall, training took over 11 hours to complete for this model.

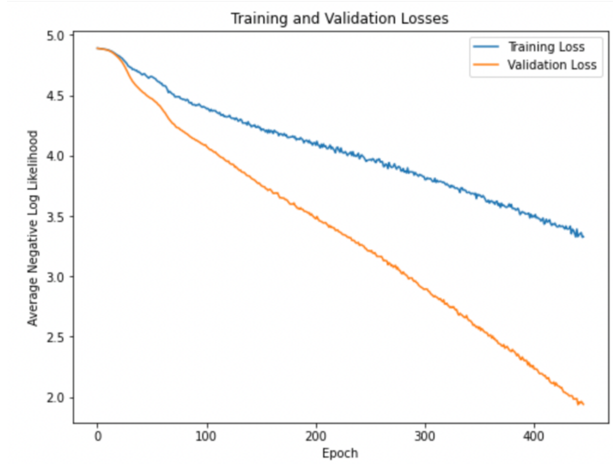

Fig. 3. Baseline Model Loss over 450 epochs

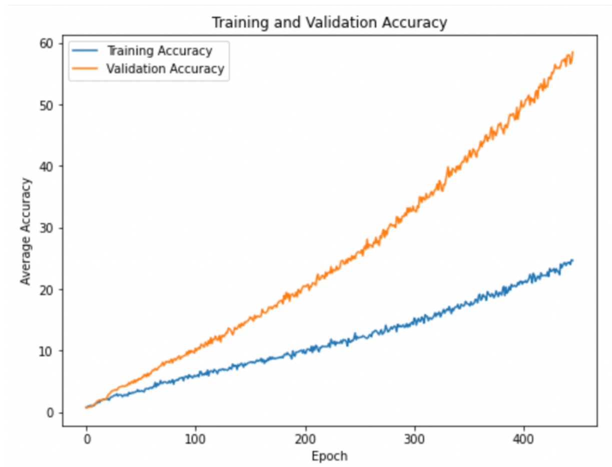

Fig. 4. Baseline Model Accuracy over 450 epochs

While this model was accurate almost 60% of the time in identifying breeds of dogs, it is only the baseline model. The goal for this model was just to have a baseline to ensure that our ResNet50 model was outperforming it.

IV. ResNet50 Model

A. Architecture

A pre-trained ResNet50 model was used. The ResNet50 model was pre-trained on ImageNet [1]. I used ResNet50 because it utilized short-cuts, or skip connections, in order to prevent over-training. This was important because there were so many learnable features for this model.

After freezing the model weights and defining the last block of the network to result in 133 output neurons, training began. Several experiments occurred in order to get the desired accuracy.

B. Experiments

Several experiments were run in order to get the accuracy for this model well above that of the baseline mode. The first four experiments were unsuccessful, and resulted in no more than 14% of accuracy for the validation data sets. However, the fifth experiment resulted in an accuracy of over 91%. The details of the experiments are shown below:

1) The last block of the network consisted of two fully-connected linear layers, one dropout layer, and one softmax layer. Early stopping after 5 epochs with no improvement and a learning rate of 1×10^{-3}. This trained for 46 epochs and resulted in no more than 10% accuracy for the validation set.
2) The last block of the network remained the same. Early stopping was changed to occur after 10 epochs with no improvement. This continued for 21 more epochs, for a total of 77 epochs, and resulted in no improvements in the model.
3) The learning rate was changed to 1×10^{-4}. After 9 more epochs, for a total of 86 epochs, the model had no more than 12% accuracy on the validation set.
4) The learning rate was changed to 1×10^{-5}. After 13 more epochs, for a total of 99 epochs, the model had no more than 15% accuracy on the validation set.
5) The last block of the network was changed to only be one fully-connected layer with 133 output neurons. The learning rate was changed back to 1×10^{-3}. The model had over 91% accuracy after trainging for 20 more epochs, for a total of 119 epochs over almost 2 hours.

C. Results

The pre-trained ResNet50 model found to be much better at classifying images of dogs into the 133 dog breeds than the baseline-model. After significantly less training, the ResNet50 model achieved an accuracy of over 91% on the validation and training sets. The accuracy and loss can be seen in Figure 5 and Figure 6.

D. Real-World Predictions

To test out how this model performs on real dogs, I tried to classify some of the dogs I know! These

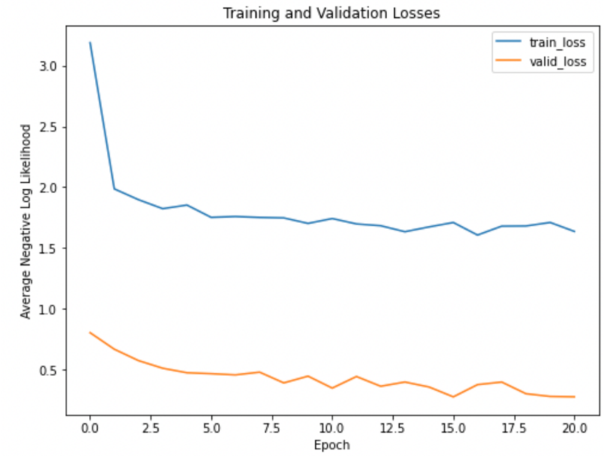

Fig. 5. ResNet50 Model Loss over the last 20 epochs

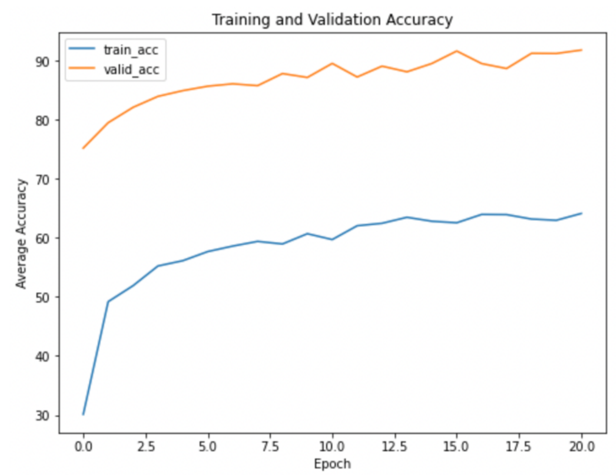

Fig. 6. ResNet50 Model Accuracy over the last 20 epochs

dogs are not of breeds which are included in this project, but the classifications make sense. In Figure 7, BeeGee, a labrodoodle (Poodle x Lab), is seen on the left. He was classified as a Bichon Frise, which can be seen on the right. Although smaller, a Bichon Frise looks very similar to the picture of BeeGee shown. In Figure 8, Major, a mutt that is presumed to be Husky x Rottweiler x Lab x Chow Chow, was classified as a Mastiff, which can be seen on the right. The similarities in coloring and head-shape make this a very believable classification. It's classifications like this one that I believe could help animal shelter staff in their adoption efforts.

V. Conclusion

In conclusion, the motivation behind this project was to create a dog breed classification system to

Fig. 7. BeeGee, a labradoodle, is seen on the left and a Bichon Frise is seen on the right

Fig. 8. Major, a mixed-breed, is seen on the left and a Mastiff is seen on the right

aid shelter staff in adoption efforts. Overall, the pre-trained ResNet50 model was able to classify images of dogs as the correct breed over 91% of the time. This accuracy carried over to real-world dog images as can be seen in Figures 7 and 8.

References

[1] Richard Socher Li-Jia Li Kai Li Jia Deng, Wei Dong and Li Fei-Fei. Imagenet: A large-scale hierarchial image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.

[2] Rebecca T. Barber Lisa M. Gunter and Clive D. L. Wynne. A canine identity crisis: Genetic breed heritage testing of shelter dogs. *PLOS ONE*, 13:1 – 16, 2018.

[3] Udacity. Dog project. [Online]. Available: https://github.com/udacity/dog-project, 2021.