Dog Face Classifier

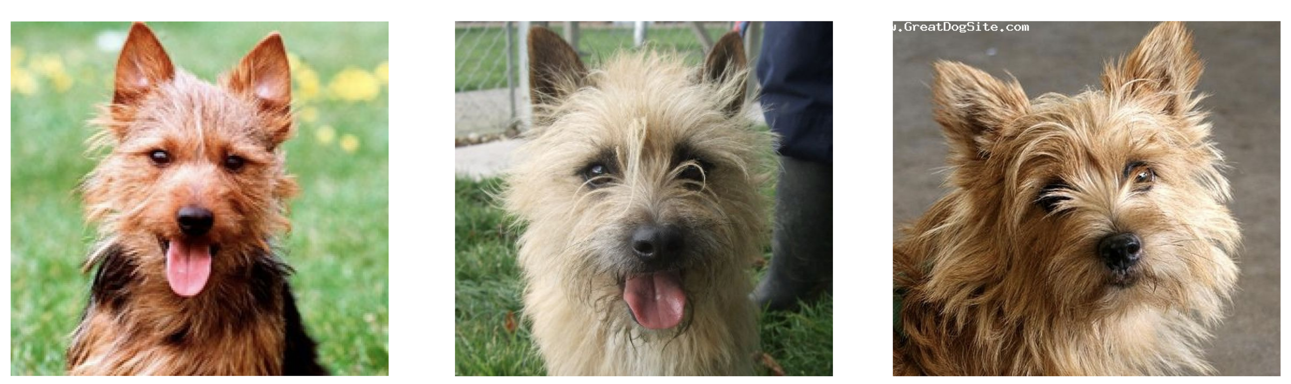
Charishma Ravoori (cr2st), Jing Sun (js6mj), Sung Min Yang (sy8pa)

University of Virginia

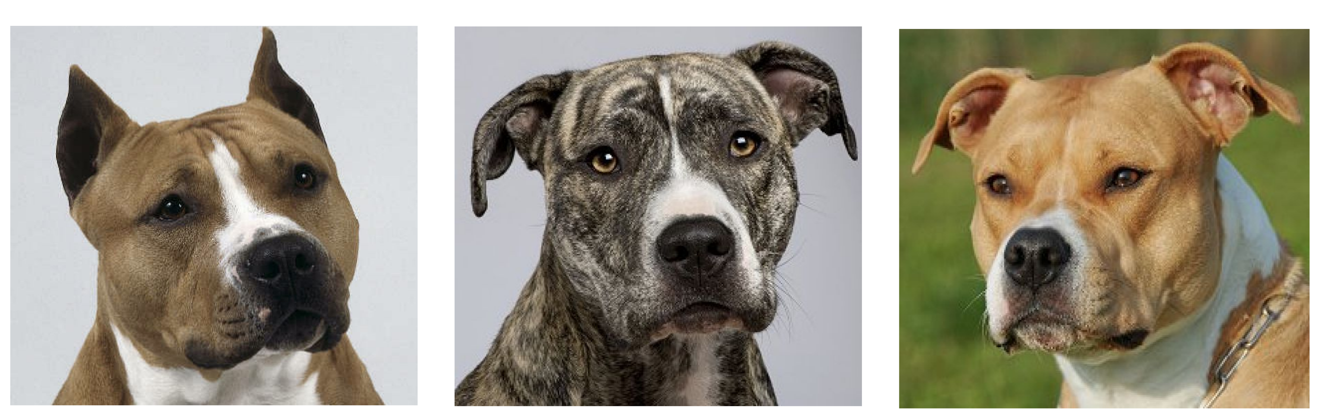
Introduction

Image classification is the task of gathering important information from image pixels and identifying the features that define what group the image belongs to. Our main objective is to identify dog breed based on the image provided. Dog breed classification is an interesting problem because dogs are both the most morphologically and genetically diverse species on Earth [1]. All breeds share similar body features and overall structure, and low inter-breed and high intra-breed variation. The algorithm's main advantage is the use of facial features extracted from the key point detection step. Dog breeds are largely identifiable from their facial features [2]. 1. larow, 2. liu

Three example images of Australian terrier, cairn terrier and Norwich terrier which display low inter-breed variation:



Three example images of American Staffordshire terrier which display high intra-breed variation:



Method

The dataset used in this study was annotated and released by Columbia researcher Jiongxin Liu [2], including 133 official breed recognized by the American Kennel Club.

Here we explored two approaches for dog breed identification:

- Random crop all images to 224x224 pixels + breed identification CNN
- Rescale all images to 256x256 pixels in grayscale + facial keypoints CNN + crop all colored images to 224x224 pixels around the identified center of the dog faces + breed identification CNN

Architecture of our facial keypoints CNN and breed identification CNN:

- Facial keypoints CNN:

Layer	Filter Size	Volume Size
Input	N/A	3 × 256 × 256
Convolution	(5, 5)	32 × 252 × 252
Max Pooling	(2, 2)	32 × 126 × 126
Convolution	(3, 3)	64 × 124 × 124
Max Pooling	(2, 2)	64 × 62 × 62
Convolution	(3, 3)	128 × 60 × 60
Max Pooling	(2, 2)	128 × 30 × 30
Convolution	(3, 3)	256 × 28 × 28
Max Pooling	(2, 2)	256 × 14 × 14
Fully Connected	N/A	50176
Fully Connected	N/A	1024
Fully Connected	N/A	16

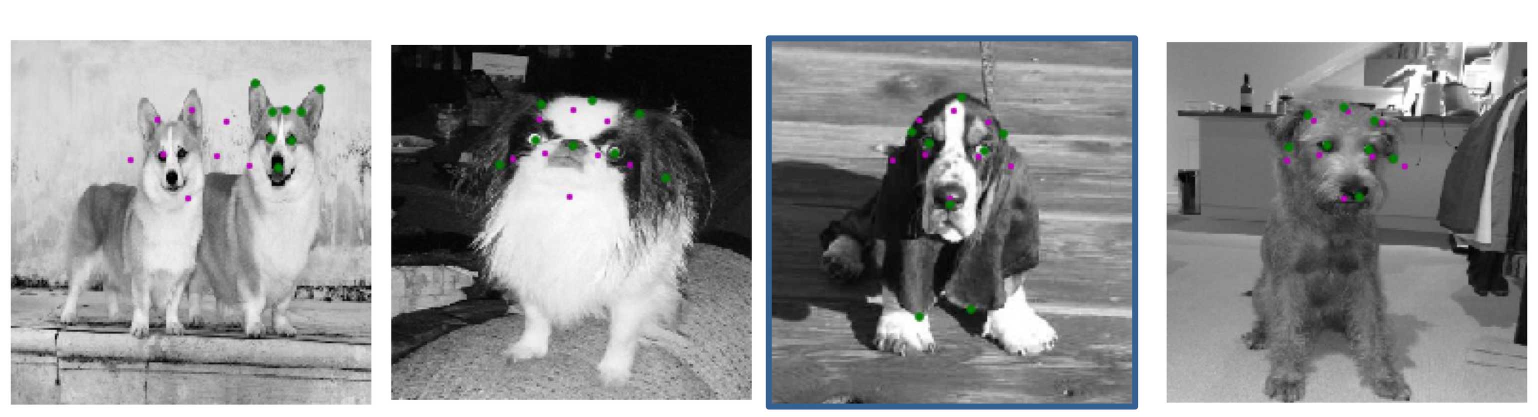
- Breed identification CNN:

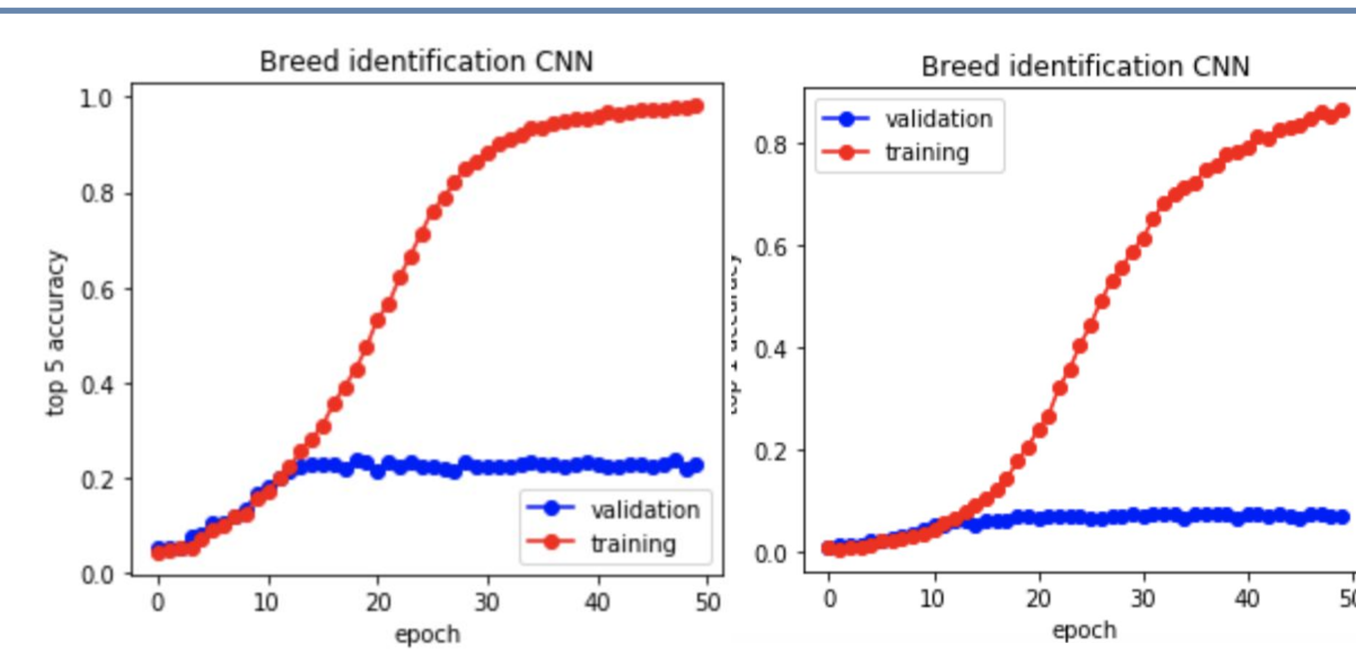
Layer	Filter Size	Volume Size
Input	N/A	(3, 224, 224)
Convolution	(5, 5)	(100, 32, 109, 109)
Max Pooling	(2, 2)	(100, 32, 54, 54)
Convolution	(3, 3)	(100, 64, 54, 54)
Max Pooling	(2, 2)	(100, 64, 27, 27)
Convolution	(3, 3)	(100, 128, 25, 25)
Max Pooling	(2, 2)	(100, 128, 12, 12)
Fully Connected	N/A	(100, 18432)
Fully Connected	N/A	(100, 200)
Softmax	N/A	(100, 133)

Experimental Results

Facial Keypoint Detection Algorithm Results

These are a sample of results from the facial keypoints detection algorithm. The green are the actual facial keypoints while the purple are our predictions. An interesting case it the picture on the left of the panel where the CNN predicts the dog on the left. However, for the most part the prediction algorithm is accurate and roughly picks up the region of the dog's face.





The plots above show the training and validation accuracies for the top 1 and top 5 images in the dataset.

Conclusions

Our facial keypoints CNN is roughly accurate in picking up facial keypoints. For the classification algorithm, it seems that performing facial cropping does improve the results, albeit just a marginal amount. For classification, retraining the pretrained Alexnet beats our CNN by a significant margin.

For future work, it would be interesting to explore other architectures to see if they have a higher success rate of detecting keypoints and for classification. Another method to try improving the accuracy is data augmentation.