Dog Face Classifier and Breed Identification

Charishma Ravoori, Jing Sun, Sung Min Yang
[cr2st, js6mj, sy8pa]@virginia.edu

Abstract

In this paper we present our project for the classification of dog breed using facial key points detection. The goal of the project is to see if using the facial key points will improve the overall prediction accuracy over just using a breed identification algorithm to classify the images. The main stages of the project are: localizing the face of a dog in the image using a convolutional neural network (CNN), crop all images centered around the facial key points of the dogs, and then feed the cropped images into another CNN for breed classification. We show that cropping with facial key points information before the classification step yields a better accuracy compared to feeding randomly cropped images through the breed identification CNN.

1. 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. In this project the main objective is to identify dog breed based on the image provided. As LaRow et al. [1] points out, dog breed classification is an interesting problem because dogs are both the most morphologically and genetically diverse species on Earth. All breeds share similar body features and overall structure, and low inter-breed and high intra-breed variation (see examples in Figure 1, 2). The algorithm's main advantage is the use of facial features extracted from the key point detection step. As Liu et al. [2] suggests, dog breeds are largely identifiable from their facial features.

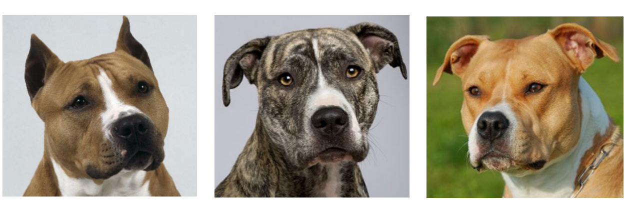

Figure 1. Three example images of American Staffordshire terrier which display high intra-breed variation.

Figure 2. Four example images of Australia terrier, Cairn terrier, Norwich terrier and Silky terrier which display low inter-breed variation.

2. Related Work

Liu et al. [2] apply appearance-based sliding window detectors on the Columbia Dogs with Parts dataset using SVM regressors with grayscale SIFT descriptors as features to detect dog faces and localize their face parts. They then train one versus all SVM models for the 133 breeds over grayscale SIFT descriptors centered at the predicted facial region, and this method yields 67% and 93% for top-1 accuracy and top-10 accuracy, respectively [2]. Rhodes [3] improves Liu's facial feature extraction algorithm by implementing a CNN trained with a mean squared error loss function on the location of the facial keypoints. The images are then rotated and scaled to a constant size for breed identification. The best network architecture, consists of six convolutional layers, three pooling layers and two fully connected layers, with implementations of dropouts and data augmentation achieves 30.6% top-1 accuracy. LaRow et al. [1] also uses CNN to identify dog facial keypoints, which are then used to extract features via SIFT descriptors and color histograms. The images are scaled to be 128 x 128 pixels, with pixel intensity values scaled to be in the range [0,1]. SVM with linear kernel achieves the best result compared to bag of words, kNN, logistic regression models with 52% top-1 accuracy, and 90% top-10 accuracy.

3. Data

The dataset used in this study is the Columbia Dogs with Parts dataset, which contains 8316 labeled images, accounting for 133 dog breeds recognized by the American Kennel Club. All images contain annotated dog facial key points: right eye, left eye, nose, right ear tip, right ear base, head top, left ear base, and left ear tip. Key point coordinates for each image are stored in a text file and released by Columbia researcher Jiongxin Liu [2].

4. Model

The methodology for dog breed identification is as follows: 1) all images are rescaled to 256x256 resolution, transformed into one channel grayscale images, then randomly cropped into 224x224 resolution; 2) a breed identification CNN was then constructed with two convolutional layers, two pooling layers, one dense layer and an output layer with 133 nodes/breeds. The architecture of the breed identification CNN is shown in Table 1.

The methodology for identifying facial key points is: 1) all images are rescaled to 256x256 resolution and matched with their corresponding key points coordinates; 2) a key points detection CNN was constructed with four convolutional layers, four pooling layers, two dense layers, and an output layer with 16 nodes for the coordinates of eight facial keypoints. The architecture of the facial key points detection CNN is shown in Table 2. After succesfully predicting the facial key points coordinates in each image, all images were cropped into 224x224 resolution centered around the center point of the faces of the dogs before feeding into the breed identification CNN for evaluation.

5. Results

The facial key points detection CNN had quite efficient results. Select sample images are shown below, where the green points are the annotated facial key points and the purple points are the predictions of those eight facial features. In most images, the algorithm was able to roughly pick up the region of the dogs' faces. However, in some interesting cases where there are multiple dogs in the same image, the facial key point detection algorithm predicted the wrong dog. The dataset were

The metrics used to examine the performance of our models were the top-1 and top-5 accuracies. The performances of our breed identification CNN with and without facial key point detection were compared against the performances of those using pre-trained AlexNet [4]. The top-1 and top-5 accuracies for each of the four methods are shown in Table 3.

Table 1. Architecture of breed identification CNN.

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

Table 2. Architecture of facial key points detection CNN.

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

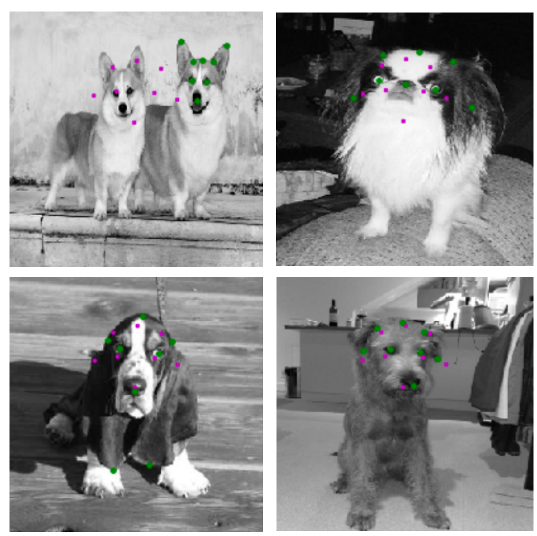

Figure 3. Four example images illustrating the facial key points detection results. Green points are the ground truth, and the purple points are our predictions.

Table 3. Top-1 and top-5 validation accuracies for our breed identification CNN and pre-trained AlexNet with and without facial key point detection. Keypoints column indicated whether key point dection was implemented to before cropping the images to 224x224 resolution.

Keypoints	**Model**	**Top-1**	**Top-5**
No	Breed CNN	6.4%	22.16%
Yes	Breed CNN	7.3%	24.06%
No	AlexNet	43.0%	75.7%
Yes	AlexNet	46.6%	77.7%

6. Conclusion

The facial key points detection CNN we constructed was roughly accurate in prediction the facial key points. For the breed identification algorithm, it seems that performing cropping around the predicted facial key points does improve the results, albeit just a marginal amount. Retraining the last layer of the pretrained AlexNet gives better performance on the breed identification 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 facial key points as well as for the breed classification. Another method that could be implemented to improve the accuracy is data augmentation.

Reference

[1] LaRow, Whitney and Mittl, Brian et al. Dog Breed Identification.

[2] Liu, Jiongxin and Kanazawa Angjoo et al. Dog Breed Classification Using Part Localization.

[3] Rhodes, Dylan. Dog Breed Identification.

[4] Krizhevsky, Alex and Sutskever, Ilya et al. ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM,* 60(6), 2017.