Dog Breed Classifier via Convolutional Neural Network

Cheng-Chieh Peng

1. Domain Background

Can a machine think and act like a human being, which can pass the Turing test so people are not able to tell is it a machine. It should possess all the abilities people can do like listening, watching, and have a meaningful response after it digests the information he receives. That's a goal for artificial intelligence which people are eager to achieve.

Computer vision is one of the important subfields. Computers record what they receive by pixels with colors as current technology. How to let the computer know what does it see, it needs to find the border, grouping pixels to a different object, and then recognize. This was thought to be extremely hard, but thanks to the recent development deep learning technique, we are making big progress on the road.

2. Problem Statement

The early training recognition tasks like handwriting numbers (MINST [1]) and the 10 classes of fashion style images (fashion-MINST[1]) in grayscale without background noise are achievable to be above 95% accuracy with shallow machine learning. People are now trying to solve more practical problems, which consist of colors and background, and also suffer the complexity from light the different angles of pictures as the problem for this project: the classification of dog breed. Different breeds of dogs might resemble in many ways, and even the same breed of dog could have different color make this problem challenge. We will build a solution based on machine learning modals to let computer output the breed from what it read.

3. Datasets and Inputs

The dataset is provided by Udacity. It consists of 133 dog breeds (each with 5-15 images) and a total of 8351 images. To utilize pre-trained deep learning Convolutional Neural Network (CNN)

models, each image is standardized for transfer learning with size 224*224 pixel, RGB color normalized to mean=[0.485, 0.456, 0.406] and standard deviation=[0.229, 0.224, 0.225].

4. Solution Statement

The input image would be recognized as it either contains dogs or not via OpenCV's implementation of Haar feature-based cascade classifiers [2]. For dog images, it is identified via either a CNN model from scratch or some other pre-trained models. The CNN model a kind of special deep learning neural networks which is composed of convolutional layers, pooling layers, and regular fully connected layers. The convolutional layer is a sophisticated way to reduce the number of parameter in deep learning (and therefore reduce the cost of computation) by applying matrix filter on layers to extract different features. In this project, 3 pre-trained models will be used : VGG [3], DenseNet [4] and ResNeXt [5].

5. Benchmark Model

The pre-trained models are trained with 1000 classes image from ImageNet with their accuracy between 60-80% [6]. The pre-trained models are available from Pytorch official site. For deep learning neural network training, the number of parameters could be a huge number (easily above millions depends on design of model), thus the computational cost for training is expensive. Using the pre-trained models, which not only the neural network structures have been determined but also the parameters have been trained to optimal value, as the foundation of transfer learning has been proven to be successfully solving classification problem while saving training cost.

6. Evaluation Metrics

The performance of models would be judged by its accuracy: the number of correctly classified dog breeds tests divided by total number of test inputs.

7. Project Design

The dog image will be either trained with CNN model from scratch or from some pre-trained models. The training will be done within limited epochs. Then comparing the results

from different models and examine which model perform better and on which breed the models fails.

[1] ‘THE MNIST DATABASE of handwritten digits’, http://yann.lecun.com/exdb/mnist/

[2] ‘Rainer Lienhart and Jochen Maydt. An extended set of haar-like features for rapid object detection.’ In Image Processing. 2002. Proceedings. 2002 International Conference on, volume 1, pages I–900. IEEE, 2002.

[3] ‘Very Deep Convolutional Networks for Large-Scale Image Recognition’, Karen Simonyan and Andrew Zisserman, arXiv:1409.1556

[4] ‘Densely Connected Convolutional Networks’, Gao Huang and Zhuang Liu and Laurens van der Maaten and Kilian Q. Weinberger, arXiv: 1608.06993

[5] ‘Aggregated Residual Transformations for Deep Neural Networks’, Saining Xie and Ross Girshick and Piotr Dollár and Zhuowen Tu and Kaiming He, arXiv:1611.05431

[6] ‘TORCHVISION.MODELS’, https://pytorch.org/docs/stable/torchvision/models.html