Capstone Project Proposal
Machine Learning Engineer Nanodegree

Aye Lu Win
September 15, 2020

Dog Breed Classification with Convolutional Neural Networks

Domain Background

Computer vision has been making rapid progress with the advances of deep learning neural networks. Convolutional Neural Networks (CNN) in particular has enabled many machine learning applications such as image classification, and object detection once considered intractable. Yann LeCun et al. introduced a 7-layer convolutional neural networks for classification of images of digits [1] in 1998. Alex Krizhevsky et al. in 2012 introduced AlexNet that outperformed existing model [2]. Karen Simonyan et al. in 2015 proposed a deep CNN known as VGG-16 that had simpler architecture [3]. A very deep CNN known as residual networks or ResNet was introduced by Kaiming He et al. [4] in 2015, and GoogLeNet [5] was introduced by C. Szegedy et al. in 2014.

The dog breed classification problem is well known supervised learning problem and has been studied by Jiongxin Liu et al., using part location [6]. It has also been studied using CNN for example by David Hsu [7] and Jesse Candido [8].

Problem Statement

The goal is to develop a CNN model for classification of dog breeds. A dog breed classification algorithm using the CNN model will also be developed. It can accept a user-supplied image and determine whether it is a dog or human. If a dog image is detected the app will predict its breed. If a human face is detected instead, the app will return the dog breed it most resembles.

Datasets and Inputs

The datasets of images for training, validation, and testing are provided by Udacity. The dog dataset has a total of 8351 images of 133 dog breeds. The dataset is divided into train, validation, and test sets with 6680, 835, and 836 images respectively. They are color images and come in different sizes and file extensions. Udacity has also provided 13233 human images of size 250 x 250 pixels. Both dog and human datasets are not balanced. Some dog breeds have more images than other breeds. The image will be preprocessed by random rotation and horizonal flip and resized to 224 x 224 pixels.

Solution Statement

Convolutional Neural Networks (CNN) will be used to classify dog breeds from the user input images. Since user can input any images, algorithms for detecting human faces and dogs are needed before classifying dog breeds. For face detection, a pre-trained face detector from OpenCV will be used. This face detector will be OpenCV's implementation of

Haar feature-based classifiers proposed by Paul Viola and Michael Jones [9]. For dog detection, a pre-trained VGG-16 model will be used. The architecture of the model was proposed by Karen Simonyan and Andrew Zisserman [3] in 2015. The weights of the model have been trained on ImageNet. For predicting dog breed, we will build a convolutional neural networks model that has convolution layers and pooling layers. We will also use a transfer learning model with ResNet50 architecture proposed by He et al. in 2015 [4].

Benchmark Model

The random chance of predicting the dog breed correctly is roughly 1 in 133 since the classes are sightly imbalanced. The accuracy for random guessing is thus less than 1%. Liu et al. in 2012 achieved 67% accuracy using part location (e.g., face and eyes) [6]. The CNN model created from scratch with three convolution layers and pooling layers will serve as the baseline model for which the targeted accuracy is greater than 10%. The goal is to achieve the accuracy greater than 80%. A transfer learning model that uses ResNet50 architecture will be used to achieve that goal.

Evaluation Metrics

Although classes are slightly imbalanced, accuracy is still a valid metrics for evaluating performance of the models and will be used as the evaluation metrics.

Project Design

The workflow of the project is as follows:

1. Import dog and human image datasets.
2. Detect human faces using OpenCV's implementation of Haar feature-base cascade classifiers and assess the performance of the face detector.
3. Detect dogs using pre-trained VGG-16 models whose weights have already been trained on ImageNet and assess the performance of the detector.
4. Preprocess train, validation and test datasets with necessary transformations such as resizing, random rotation, and converting to tensors.
5. Create a CNN model from scratch with three convolution layers and pooling layers, train the model and evaluate its performance.
6. Create a ResNet50 model using transfer learning, train the model and evaluate its performance.
7. Write a dog breed classification algorithm that accepts a user input image and determines whether it contains a human, dog or neither. If a dog is detected, its predicted breed is returned. If human face is detected, the dog breed it most resembles is returned. If neither is detected, and error message is returned.
8. Test the dog breed classification algorithm with different user input images.

References

[1]. Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner, "Gradient-Based Learning Applied to Document Recognition", Proceedings of the IEEE, Vol. 86 (1998).

[2]. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Network", NIPS (2012).

[3]. Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition", ICLR (2015).

[4]. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep Residual Learning for Image Recognition", https://arxiv.org/abs/1512.03385, (2015).

[5]. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going Deeper with Convolutions", arXiv:1409.4842, (2014).

[6]. Jiongxin Liu, Angjoo Kanazawa, David Jacobs, and Peter Belhumeur, "Dog Breed Classification Using Part Location", European Conference on Computer Vision, Springer, Berlin, Heidelberg, 2012.

[7] David Hsu, "Using Convolutional Neural Networks to Classify Dog Breed", http://cs231n.stanford.edu/reports/2015/pdfs/fcdh_FinalReport.pdf (2015).

[8] Jesse Candido, "Dog Breed Classification and Visualization", http://cs230.stanford.edu/projects_spring_2018/reports/8291007.pdf (2018).

[9]. Paul Viola and Michael Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", Conference on Computer Vision and Pattern Recognition (2001).