Dog Breed Classifier using Convolutional Neural Network

Reza Jalayer

January 2021

Domain Background

Image classification has a wide range of real-world applications and is used in different fields including automotive industry, security industry, health care industry, gaming and social media platforms. It also plays an important role in remote sensing images and is used for various applications such as environmental change, surveillance, geographic mapping, disaster control [1]. Correct classification has vital importance, especially in medicine for example. Therefore, improved methods are needed in this field.

Deep learning [2] is a powerful machine learning technique for solving a wide range of computer applications including image classification. It is composed of multiple processing layers that can learn more powerful feature representations of data with multiple levels of abstraction. In the deep learning technique, several numbers of models are available such as convolutional neural network (CNN).

Deep convolutional neural networks [3] provide better results than existing methods in the literature due to advantages such as processing by extracting hidden features, allowing parallel processing and real time operation. The concept of convolutions in the context of neural networks begins with the idea of layers consisting of neurons with a local receptive field, i.e., neurons which connect to a limited region of the input data and not the whole [4].

Nearly every year since 2012 has given us big breakthroughs in developing deep learning models for the task of image classification. Due to its large scale and challenging data, the ImageNet (image-net.org) challenge has been the main benchmark for measuring progress. some of the major architectures that made that progress possible are:

AlexNet: The 2012 paper from the university of Toronto became one of the most influential papers in the field after achieving a nearly 50% reduction in the error rate in the ImageNet challenge [5].

VGGNet: The 2014 Oxford University paper extended the idea of using a deep networking with many convolutions and ReLUs (Rectified Linear Units). A deep network with lots of small 3x3 convolutions and non-linearities [6].

Problem Statement

In this project I will go through building a machine learning model for dog breed classification. The problem starts by first distinguishing an image as a human or a dog image, and if it is detected as a dog image the dog breed will be predicted. If a human image is detected, then a “resembling” dog breed will be the shown output.

Datasets and inputs

The dataset used for this project includes 13233 human images as well as 8351 dog images. The dog images are grouped into 133 different breeds (classes). The images are RGB images and the average size of a dog image is 529by567 pixels. Each human image contains at least one human (face) and each dog image contains only one dog.

The human/dog images will be used to detect a human or dog, and then the dog images will be used to train a dog breed classifier.

Dog images data is an unbalanced dataset, number of samples in each of the 133 classes shown below and it is clear that the dataset is unbalanced (i.e. different number of samples per class)

```
array([96, 93, 92, 89, 87, 86, 86, 84, 83, 83, 82, 82, 82, 81, 81, 81, 80,
       80, 80, 80, 80, 80, 80, 79, 79, 79, 79, 78, 78, 78, 77, 77, 76, 76,
       75, 74, 74, 73, 73, 72, 71, 71, 71, 70, 70, 68, 67, 67, 66, 66, 66,
       66, 66, 66, 66, 66, 65, 65, 64, 64, 63, 63, 63, 63, 63, 62, 62, 62,
       62, 62, 62, 62, 62, 61, 61, 60, 60, 60, 60, 60, 59, 59, 59, 59, 58,
       58, 58, 57, 57, 56, 56, 55, 55, 55, 55, 55, 55, 54, 54, 53, 53, 53,
       52, 51, 51, 51, 50, 49, 49, 49, 46, 46, 44, 44, 44, 42, 42, 42, 42,
       41, 41, 40, 39, 39, 38, 38, 38, 37, 37, 36, 35, 33, 33])
```

Solution Statement

As mentioned in the domain background section, CNN is a good technique for such image classification problem. The CNN training/validation input will be the RGB dog images. In this project two CNN models will be used: CNN from scratch (that can serve as a benchmark) and CNN using transfer learning from one of the state of the arts model mentioned in domain background section (i.e. VGG16)

Benchmark model

A simple benchmark model that will be used is multi-class classification logistic regression (using sk-learn package) just to show the basic feasibility of the dog breed classification using the provided input data. We do not expect a good accuracy for this benchmark model since this may not be a good approach for such complicated image analysis.

The next model that can also act as a benchmark will be a CNN model designed from scratch. It would be something like a 2-3 convolutional/maxpool layers followed by 1-2 fully connected layers. We expect a higher result accuracy compared to the first benchmark.

Evaluation Metrics

Cross Entropy loss (log loss for multi class data) will be calculated during model training as well as model testing on test data (10% data of the total dog images).

$$CE_{loss} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} t_{ij} \log\left(p_{ij}\right)$$

Where N and C are number of samples and classes, t_{ij} is only 1 if sample i belongs to class j, *and* p_{ij} is the predicted probability.

Also the accuracy will be tested on test data by simply calculating number of “correct predictions” divided by “total predictions”. However this is an unbalanced dataset, so this accuracy may not be a good indicator because classes with fewer sample would be affected by classes with more data samples.

Project Design

The summary of the project structure is as follow:

- Implement logistic regression benchmark model using sk-learn (benchmark)
- Implement a dog detector and a human (face) detector logic
- Design a CNN from scratch, train and test it with the dog image dataset and measure the accuracy
- Use a pretrained VG16 CNN, train (only the last fc layer) and test with the dog image dataset, and measure the improved accuracy
- Use the dog/human detector of step 2 and the transferred CNN of step 4 to implement the final solution: if a dog image is detected, predict the breed; if a human is detected, show the resembling breed and if no human or dog detected, show some error message.

References

[1] P. Deepan, L.R. Sudha, The Cognitive Approach in Cloud Computing and Internet of Things Technologies for Surveillance Tracking Systems, 2020

[2] Yann LeCun, Yoshua Bengio and Geoffrey Hinton, Deep Learning, Nature, VOL 521, MAY 2015

[3] Saad Albawi, Tareq Abed Mohammed, Saad Al-Zawi, Understanding of a convolutional neural network, International Conference on Engineering and Technology (ICET), Aug 2017

[4] ralhad Gavali ME, J. Saira Banu PhD, Deep Learning and Parallel Computing Environment for Bioengineering Systems, 2019

[5] Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, NIPS, 2012

[6] Karen Simonyan and Andrew Zisserman, VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION, In Proc. ICLR 2015