ADENIYI ADEBOYE (MACHINE LEARNING NANODEGREE)

DOG BREED IMAGE CLASSIFICATION PROJECT'S PROPOSAL

1.0 BACKGROUND

Computer vision as a whole encompasses different categories which include image classification, localization, segmentation and object detection. Among those, image classification has been one of the most fundamental and most sort-after aspect or problem of this field, simply because of its application in recognizing and classifying a wide variety of objects and images. As the name implies, image classification can be defined as the identification and classification of an image based on a given group of image classes. This entails, given an image, identify it as being a member of one of several fixed classes.

Image classification has numerous applications in various fields and industries such as in autonomous driving, where fast image classification is required for switching lanes, social media platforms such as Facebook, use image classification to personalize and improve user experience on those platforms. In this project, the type of image classification that we would be concern with is called Fine-grained image classification. Fine grained image classification targets at distinguishing fine-level image categories in images i.e. classifying images with high interclass similarity and large intraclass variations such as bird species, dog breeds and airplane types to mention but a few.

Traditional machine learning approaches was utilized in the nineties to classify images, but this yielded very low accuracies and were faced with several challenges such as hand-crafted feature extraction process. However, in recent years deep neural networks (DNN) also termed Deep learning has found complex formation and underlying patterns in large training dataset to be able to classify images correctly. Amongst these DNN are Convoluted Neural Networks (CNN or ConvNET) which have demonstrated phenomenal achievements in computer vision problems, especially in image classification.

CNN is a type of multi-layer neural network inspired by the mechanism of the optical system of living creatures. CNN architectures make assumption that the inputs are images (they take into account the spatial dimension of the input), which allows us to encode certain properties into the architecture. A typical CNN is composed of a single or multiple blocks of a convoluted layers and sub-sampling or pooling layers, thereafter, one or more fully connected layers and finally an output layer (figure 1).

2.0 MOTIVATION AND PROBLEM STATEMENT

Diverse breeds of dogs exist in today's world and due to their ubiquity they are not easy or extremely difficult to identify, for example, someone may want to acquire a new dog he/she has seen before, or maybe someone is looking for his/her missing dog, or people may wants to know the name, features, species or attributes of a beautiful/agile/elegant dog walking down the streets in real time. All of these problems could be solved by just uploading the picture of the dog of interest on an app, which classify the dog of interest under a certain category or class and further display its name, family type, attributes and other relevant information about the dog.

To be able to create or develop such an app, you will need a deep learning model that does all the background work of dog breed classification and thereafter outputs the correct class of the dog. This aforementioned scenario motivates the concept of dog breed classification which could be accomplished using CNN.

In other words, the aim of this project is to build a fine-grained image classification model which classifies dog breeds given a specific number of different dog classes. The problem being investigated in this project comes with classification of dog breeds correctly using CNN (by building the model from scratch and through transfer learning).

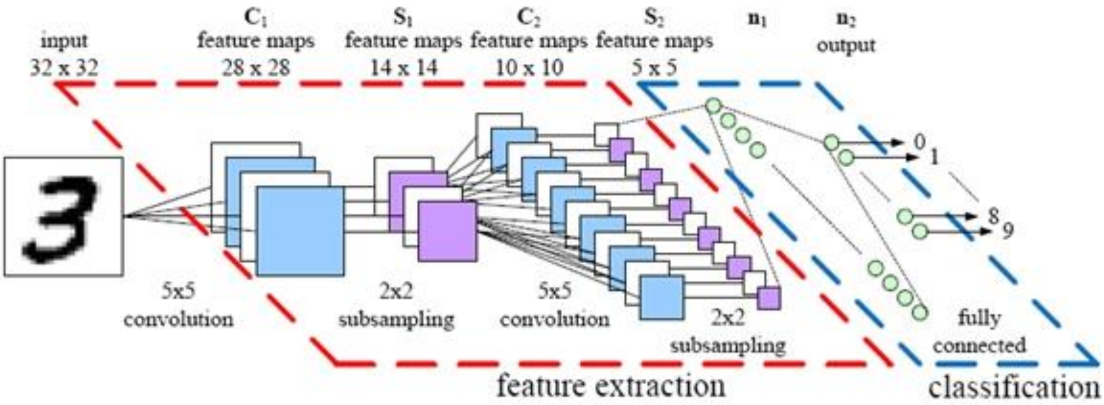

Figure 1: *A convoluted neural network built for a handwritten digit classification. From the left: we have the handwritten image(3), convoluted layer (with four feature maps), pooling layer(four feature maps), another convoluted layer (with 12 feature maps), another pooling layer (with 12 feature maps), these convoluted and pooling layers are basically used for feature extraction. followed by them, fully connected layer i.e. each neuron in this layer is connected to all activation values from the previous layer, then the output layer that has the number of classes that we want to predict. Note that the filter size for the convoluted layers is 5 x 5, max pooling filter is 2 x 2.*

3.0 DATASETS AND INPUT

The dataset to be utilized for this project will be acquired from Scikit-Learn package. The dog dataset consists of 133 dog breed classes with a total of 8351 total dog breeds. On an average, we have between 50 – 65 dog images per class. This suggest that the dog dataset is relatively balanced. The individual dog images are of different pixels i.e. mostly between 300 – 550 pixels. These dog images will be resized to 224 x 224 pixels during the data processing step as preparation for building the CNN model.

Furthermore, another dataset that contain human faces is also provided by Udacity, the dataset comprises of 13234 images and this will also be assessed along with the dog dataset basically to detect if an image is a dog, human or neither of the two.

4.0 SOLUTION STATEMENT

As highlighted in section 2.0, the aim of this project is to build a fine-grained dog breed classification model using CNN. This model will be built from scratch and also with transfer learning.

First of all, the model will be built using the CNN architecture which basically consist of Convolutional layers, Pooling layers, Fully connected layers and an output layer (figure 1). To be more elaborate, CNNs have two types of neural layers, first is the convolutional layer – this layer runs a sliding window (local receptive field) through the input image and at each step convolves this subimage with a filter, producing a Feature map (also known as Depth slice), once the filter weights (a filter weight(s) & bias is shared among the neurons in a feature map) changes a new feature map is created, a convoluted layer / volume consist a number of filters that correspond to numbers of feature maps or depth slices.

A pooling layer downsamples the convoluted layer along the spatial dimensions by a certain factor. This reduces the size of the representation and the number of parameters which makes the model more computationally efficient and prevents overfitting. The common type of pooling layer is max-pooling layer, which takes the max from each downsampled block.

The fully connected layers consist of neurons, where each of them have full connections to all activations from the previous layer as seen in regular or artificial neural networks.

The output layer consist of the different classes of images such as different dog breeds classes that will be used to classify the incoming image.

The hyperparameters for this CNN will be tweaked and tuned appropriately with the hope that best hyperparameters will be obtained to get accurate predictions and high accuracy scores.

Furthermore, the model will also be built through transfer learning. In this context, transfer learning means to apply or transfer the knowledge that a deep learning model holds (as represented by its parameters) to a new (some way related) task (which here is our dog breed classification task). Fine tuning will be done to some of the layers embedded in this pretrained DeepCNN, through some fine adjustments to further optimize the performance of this pretrained model on our specified task.

In most cases, including our dog breed classification task, the first few layers in the pre-trained DCNN to be used for transfer learning will not be fine-tuned (i.e. left frozen), this is because this first few layers learn the low level generic features such as edges and corners which are generic or common to most image classification tasks and datasets while the top layers (i.e. next-to-last layer) of the convolutional neural network which corresponds to the major classification step would need to be completely replaced with another classifier for the task that is to be accomplished, this is because if not replaced the categories this pretrained DCNN will output will not correspond to our desired product image classification at all. Other layers in between the first few layers and the next-to-last layers will undergo fine-tuning (which could be done through using smaller learning rate for these layers assuming that the pre-trained DCNN weights are effective).

5.0 BENCHMARK MODEL

The benchmark model that would be used for comparison to our optimized model would be *a CNN model built from scratch by us*. This model will further ascertain that the dog breed classification problem is solvable.

We came across other research papers and projects that used various state-of-art image classification models for dog breed classification problems, for example, Devikar, 2016 used Inception v3 as the pretrained model and achieved 96% accuracy for 11 dog classes (275 non-popular dog dataset), Higa, 2019 used VGG-16 and DenseNet-201 as the pre-trained model and achieved 82.71% accuracy for VGG-16 and 89.33% accuracy for DenseNet-201 (Stanford dog dataset), Chen et al, used VGG-19 for transfer learning and achieved 68% for 120 dog class classification (Stanford dog dataset).

However, we will not use any of the aforementioned models as benchmarks for our problem, this is simply because the models were trained on completely different dog datasets other than ours.

6.0 EVALUATION METRICS

The measure of our model performance will be based on the model's accuracy score (in percentage). The true measure of the model's performance is based on the test accuracy, which represent the trained model classifying on completely (never seen before) new images.

While training, the loss is also a vital parameter to watch out for, if it is decreasing incrementally during training then the model is actually learning from the training data but if otherwise then you should consider monitoring the model closely, and if further persist, you should stop the model.

7.0 PROJECT DESIGN

The project will be executed by following the steps already outline in the dog breed classification breed jupyter notebook provided by Udacity, and it includes:

1. Importation of the Dog dataset

2. Humans would be detected using OpenCV Haars feature-based cascade classifiers, this will be useful towards the end of the project to detect whether the image is either a human or a dog.

3. Data pre-processing. Here, the dog images would be resized to the required pixels that they can be trained on and some other additional pre-processing tasks will be done to prepare the data.

4. Creation of CNN from scratch to classify dog breeds.

5. Use the CNN created to classify the dog breeds.

6. Create CNN (with transfer learning) to classify the dog breeds i.e. using a pre-trained model that is going to be fine-tuned, for the dog breed classification.

7. Write an algorithm that depicts whether the input image is a dog, human or neither of the two.

References

C. Szegedy *et al.*, "Going deeper with convolutions," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, 2015, pp. 1-9. doi: 10.1109/CVPR.2015.7298594

Xavier S. Higa, 2019. Dog Breed Classification Using Convolutional Neural Networks: Interpreted Through a Lockean Perspective. **Lake Forest College Publications.** Senior Theses Student Publications, 4-7-2019. https://publications.lakeforest.edu/seniortheses

He, Kaiming, Zhang, Xiangyu, Ren, Shaoqing, and Sun, Jian. Delving deep into recti_ers: Surpassing human-level performance on imagenet classi_cation. In Proceedings of the IEEE International Conference on Computer Vision, pp. 1026{1034, 2015a.

He, Kaiming, Zhang, Xiangyu, Ren, Shaoqing, and Sun, Jian. Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385, 2015b.

Pratik Devikar, 2016. Transfer Learning for Image Classification of various dog breeds. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5, Issue 12, December 2016. ISSN: 2278 – 1323*

Farhana Sultana, Abu Sufian, Paramartha Dutta, 2018. Advancements in Image Classification using Convolutional Neural Network. *This paper has been accepted and presented on 2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks. This is preprint version and original proceeding will be published in IEEE Xplore.* 978-1-5386-7638-7/18/$31.00 © 2018 IEEE

Yizhou Chen, Xiaotong Chen, Xuanzhen Xu. Dog Breed Classification via Convolutional Neural Network

https://mc.ai/dog-breed-identification-using-deep-learning/