**Dog Breed Identification-Determine the breed of a dog in an image**

***Debajyoti Das****, AIML Student****, Varun Sivasubramanian****, AIML Student,* ***Aranya Das****, AIML Student*

**Abstract:** In this paper, we analyze the different possible methods to correctly predict the breed of a dog based on an image from the available dataset. The idea is to design & develop a computer vision model capable of correctly classifying the breed from a dog image. We have tried to develop & design this solution by using various pre-trained models for feature extraction & then using a custom computer vision solution to correctly classify the image based on the aforesaid features.

Additionally, we have also used Image Augmentation techniques to create a greater number of images for both training & testing and check if the model performance, both my means of prediction accuracy, as well as by computational resources requirements, can be improved.

The dataset is part of a Kaggle Playground competition, wherein we were provided with a training set and a test set of images of dogs. Each image has a filename that is its unique id. The dataset comprises 120 breeds of dogs. The goal of the competition was to create a classifier capable of determining a dog's breed from a photo.

We will utilize various pre-trained models like Resnet50, VGG16, and even a custom CNN built from scratch for achieving our goal. While the pre-trained models can be used for feature-extraction, the custom solution will be trained for both feature extraction & classification. We will also experiment by concatenating the extracted features from multiple pre-trained models to finally feed a basic DNN for classification. The model performance will be judged using Multi Class Log Loss between the predicted probability and the observed target.

We will also utilize different call-back techniques like **Checkpoint**, **EarlyStopping**, and **ReduceLROnPlateau** for efficient utilization of available computational resources without sacrificing on model accuracy. This will also help us in saving the model training weights & probably help in building a model capable of identifying breeds of other animals, or probably ID’ing different dogs which can be built as a solution to track missing dogs. As bounding boxes are available as annotations, we can also expand our analysis to Image Classification+ Localization.

1. **Introduction**

Kaggle is an online community of data scientists and machine learning practitioners. It started with the intention of allowing users to find and publish data sets, explore and build models in a web-based data-science environment, work with other data scientists and machine learning engineers, and enter competitions to solve data science challenges. It started in 2010 by offering machine learning competitions. Today it also offers a public data platform, a cloud-based workbench for data science, and Artificial Intelligence education. In March 2017; Fei-Fei Li, Chief Scientist at Google, announced Google’s intention of acquiring Kaggle during her keynote at Google Next & from March 8, 2017 Kaggle is a subsidiary of Google LLC.

Kaggle competitions regularly attract over a thousand teams and individuals. Kaggle's community has thousands of public datasets and code snippets (called "kernels" on Kaggle). Many of these participants are researchers who then go on to publish papers in peer-reviewed journals based on their performance in Kaggle competitions. Kaggle’s Machine learning competitions were the first of Kaggle's product. In these competitions companies posted problems and machine learners competed to build the best algorithm. Kaggle has run hundreds of machine learning competitions since the company was founded. These competitions range from improving gesture recognition for Microsoft Kinect to making an football AI for Manchester City to improving the search for the Higgs boson at CERN. Competitions have resulted in many successful projects including furthering the state of the art in HIV research, chess ratings and traffic forecasting. Several academic papers have been published on the basis of findings made in Kaggle competitions. A key to this is the effect of the live leader board, which encourages participants to continue innovating beyond existing best practice.

The Public datasets platform is another popular offering from Kaggle wherein community members share datasets with each other. These datasets cover everything from bone x-rays to results from boxing bouts. The Dog Breed Identification is a Kaggle Playground Competition which utilises the Stanford Dogs Dataset for this competition. The Stanford Dogs dataset contains images of 120 breeds of dogs from around the world. This dataset has been built using images and annotation from ImageNet for the task of fine-grained image categorization. There are 120 categories in this dataset, covering 20,580 images. The images annotated for bounding boxes and class labels. The dataset was created by Aditya Khosla, Nityananda Jayadevaprakash, Bangpeng Yao, and Fei-Fei Li.

The objective of this paper is to design & develop a computer vision solution that is capable of correctly classifying an image from the aforementioned dataset into its associated breed. The Kaggle competition evaluated the submissions using Multi-Class Log Loss between the predicted probability and the observed target. We will use similar metrics, along with standard metrics like accuracy. Additionally the Stanford dataset proposes 2 baseline metrics: Mean Accuracy-wherein the accuracy per class is measured as the number of images varies from 1-100, Comparison of Accuracy per Class- The accuracy of each class is compared for 15 and 100 training images per class. As bounding boxes are available as annotations, we can also expand our analysis to Image Classification+ Localization.

1. **Tentative List of Algorithms :**

Deep Learning (DL) (Bengio, 2009) [1]is a branch of Artiﬁcial Intelligence that aims at developing techniques that allow computers to learn complex perception tasks, such as seeing and hearing, at human level of accuracy. It provides near-human level accuracy in image classiﬁcation, object detection, speech recognition, natural language processing, vehicle and pedestrian detection, and more.

A Deep Learning approach particularly eﬀective for vision tasks exploits Convolutional Neural Net- works (CNN) (Krizhevsky et al., 2012; Simonyan & Zisserman, 2014; Girshick et al., 2014)**[2].** A CNN is composed of a possibly large number of hidden layers, each of which performs mathematical computations on the input provided by the previous layer and produces an output that is given in input to the following layer. A CNN diﬀers from classical neural networks for the presence of convolutional layers, which can better model and discern the spatial correlation of neighbouring pixels than normal fully connected layers.

For a classiﬁcation problem, the ﬁnal outputs of the CNN are the classes which the network has been trained on. The training phase is usually extremely expensive from a computational point of view, and may take a long time to complete. Once the network has been trained and the classiﬁer has been initialized accordingly, the run time phase of prediction is quite fast and eﬃcient. In case of Image Classification + Localization, we have different approaches like Single-Shot Detection or Region Proposal based algorithms like YOLO & SSD for the former and R-CNN, Faster R-CNN for the latter. Region-Proposal based algorithms often time use another fundamental concept called Image Segmentation for running algorithms like Selective Search which produce something called Regions of Interest. Other than that these algorithms also use CNNs as the fundamental building blocks for their architecture.

Some examples of CNN that are planned to be implemented in our project are:

1. ***ALEXNET*:** A popular model using CNN algorithm for this kind of image classification is Alexnet. The architecture of an AlexNet consists of 60 million parameters and 500,000 neurons. It is organized into ﬁve convolutional layers, some followed by max- pooling layers, and two fully connected layers with a 1000-way softmax (more details can be found in (Krizhevsky et al., 2012))[2]. It consists of 11x11, 5x5,3x3, convolutions, max pooling, dropout, data augmentation, ReLU activations, SGD with momentum. It attached ReLU activations after every convolutional and fully-connected layer. AlexNet was trained for 6 days simultaneously on two Nvidia Geforce GTX 580 GPUs which is the reason for why their network is split into two pipelines.
2. ***VGG16*:** VGG16 is a convolution neural net (CNN ) architecture which was used to win ILSVR(Imagenet) competition in 2014. It is considered to be one of the excellent vision model architecture till date. Most unique thing about VGG16 is that instead of having a large number of hyper-parameter they focused on having convolution layers of 3x3 filter with a stride 1 and always used same padding and maxpool layer of 2x2 filter of stride 2. It follows this arrangement of convolution and max pool layers consistently throughout the whole architecture. In the end it has 2 FC(fully connected layers) followed by a softmax for output. The 16 in VGG16 refers to it has 16 layers that have weights. This network is a pretty large network and it has about 138 million (approx.) parameters.
3. ***RESNET [4]*:** The major drawback about the **VGG16 , Alexnet** wasthat they were very deep NN and werehardto trainbecause of the notorious vanishing gradient problem — as the gradient is back-propagated to earlier layers, repeated multiplication may make the gradient infinitively small. As a result, as the network goes deeper, its performance gets saturated or even starts degrading rapidly. The core idea of ResNet is introducing a so-called “identity shortcut connection” that skips one or more layers. The Skip Connections between layers add the outputs from previous layers to the outputs of stacked layers. This results in the ability to train much deeper networks than what was previously possible.

**References**

1. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it>
2. ["Novel Coronavirus—China"](https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/). [*World Health Organization*](https://en.wikipedia.org/wiki/World_Health_Organization) *(WHO)*. Retrieved 9 April 2020
3. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. (February 2020). ["Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China"](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7159299). *Lancet*. **395** (10223): 497–506. [doi](https://en.wikipedia.org/wiki/Doi_(identifier)):[10.1016/s0140-6736(20)30183-5](https://doi.org/10.1016%2Fs0140-6736%2820%2930183-5). [PMC](https://en.wikipedia.org/wiki/PMC_(identifier)) [7159299](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7159299). [PMID](https://en.wikipedia.org/wiki/PMID_(identifier)) [31986264](https://pubmed.ncbi.nlm.nih.gov/31986264)
4. ["Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV)"](https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)). [*World Health Organization*](https://en.wikipedia.org/wiki/World_Health_Organization) *(WHO)*. 30 January 2020. [Archived](https://web.archive.org/web/20200131005904/https:/www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)) from the original on 31 January 2020. Retrieved 30 January 2020
5. ["WHO Director-General's opening remarks at the media briefing on COVID-19—11 March 2020"](https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020). [World Health Organization](https://en.wikipedia.org/wiki/World_Health_Organization). 11 March 2020. Retrieved 11 March2020.
6. ["COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)"](https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6). [*ArcGIS*](https://en.wikipedia.org/wiki/ArcGIS). [Johns Hopkins University](https://en.wikipedia.org/wiki/Johns_Hopkins_University). Retrieved 14 June 2020.
7. ["Ergänzung zum Nationalen Pandemieplan – COVID-19 – neuartige Coronaviruserkrankung"](https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Ergaenzung_Pandemieplan_Covid.pdf?__blob=publicationFile) (PDF). *rki.de*. Robert Koch Institute. Retrieved 4 March 2020.
8. de Winter, J.C.F. (2013) "Using the Student's t-test with extremely small sample sizes," Practical Assessment, Research, and Evaluation: Vol. 18, Article 10.
9. T test as a parametric statistic: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4667138/>. Retrieved on 14 June 2020
10. Hypothesis Testing - Analysis of Variance (ANOVA): Lisa Sullivan, PhD. Professor of Biostatistics. Boston University School of Public Health
11. DYNAMICS OF MEASLES EPIDEMICS: ESTIMATING SCALING OF TRANSMISSION RATES USING A TIME SERIES SIR MODEL: [Ottar N. Bjørnstad](https://esajournals.onlinelibrary.wiley.com/action/doSearch?ContribAuthorStored=Bj%C3%B8rnstad%2C+Ottar+N), [Bärbel F. Finkenstädt](https://esajournals.onlinelibrary.wiley.com/action/doSearch?ContribAuthorStored=Finkenst%C3%A4dt%2C+B%C3%A4rbel+F), [Bryan T. Grenfell](https://esajournals.onlinelibrary.wiley.com/action/doSearch?ContribAuthorStored=Grenfell%2C+Bryan+T). First published:01 May 2002.