# Capstone Proposal

Pankaj Patil
January 10, 2021

## Domain Background

*ImageNet is a huge dataset of labelled images and is widely used for comparing current benchmarks in the field of computer vision. ILSVR competition involves training a model on 1.2 million training images, 50k validation images and 10k test images to predict correct category for an image among 1000 available categories. The target classes include various objects from day-to-day lives like dogs, cats, various household projects, vehicle types etc.*

*In 2014, the 1st runner up for the ILSVR competition was VGGNet. VGGNet is invented by Visual Geometry Group (by Oxford University). The reason to understand VGGNet is that many modern image classification models are built on top of this architecture. It is a convolutional neural network made of architecture shown in column D of image below:*

| ConvNet Configuration | | | | | |
|---|---|---|---|---|---|
| A | A-LRN | B | C | D | E |
| 11 weight layers | 11 weight layers | 13 weight layers | 16 weight layers | 16 weight layers | 19 weight layers |
| input (224 × 224 RGB image) | | | | | |
| conv3-64 | conv3-64<br>**LRN** | conv3-64<br>**conv3-64** | conv3-64<br>conv3-64 | conv3-64<br>conv3-64 | conv3-64<br>conv3-64 |
| maxpool | | | | | |
| conv3-128 | conv3-128 | conv3-128<br>**conv3-128** | conv3-128<br>conv3-128 | conv3-128<br>conv3-128 | conv3-128<br>conv3-128 |
| maxpool | | | | | |
| conv3-256<br>conv3-256 | conv3-256<br>conv3-256 | conv3-256<br>conv3-256 | conv3-256<br>conv3-256<br>**conv1-256** | conv3-256<br>conv3-256<br>**conv3-256** | conv3-256<br>conv3-256<br>conv3-256<br>**conv3-256** |
| maxpool | | | | | |
| conv3-512<br>conv3-512 | conv3-512<br>conv3-512 | conv3-512<br>conv3-512 | conv3-512<br>conv3-512<br>**conv1-512** | conv3-512<br>conv3-512<br>**conv3-512** | conv3-512<br>conv3-512<br>conv3-512<br>**conv3-512** |
| maxpool | | | | | |
| conv3-512<br>conv3-512 | conv3-512<br>conv3-512 | conv3-512<br>conv3-512 | conv3-512<br>conv3-512<br>**conv1-512** | conv3-512<br>conv3-512<br>**conv3-512** | conv3-512<br>conv3-512<br>conv3-512<br>**conv3-512** |
| maxpool | | | | | |
| FC-4096 | | | | | |
| FC-4096 | | | | | |
| FC-1000 | | | | | |
| soft-max | | | | | |

*Image Source : Table 1 of Very Deep Convolutional Networks for Large Scale Image Recognition, Simonyan and Zisserman (2014).*

*Authors of VGGNet knew that deeper neural networks struggle with convergence. So they trained shallower networks and used weights from them to initialize the deeper networks. While better models are now available, VGGNet still remains one of the best known algorithms in image classification tasks. I plan to use it for solving my problem statement.*

## Problem Statement

The problem statement I propose to solve is for a given input image, developing a model to classify between human and dogs. Then if a dog, the model would further classify the breed of the dog. The user would be able to upload an image via a webapp and then the webapp would communicate with an endpoint for inference. This would involve use of pre-trained state-of-the-art like VGGnet to do classification tasks.

## Datasets and Inputs

There would be 2 datasets for this project:

a) Dog Dataset: This would contain RGB images of various dogs and a dataset specifying which image corresponds to which breed. The dataset I'll be using has 8,351 dog images and has been provided by Udacity
b) Human Dataset: Same as dog dataset but it would contain images of humans. Since all images would correspond to humans, there would be no labels required. The dataset I'll be using has 13,233 human images and has been provided by Udacity

Since the task at hand is to first predict if a human/dog is present in the give image, it is important that we have a model which is able to classify either. Thus, we require dataset which has images of both humans and dogs. I plan to create a CNN to identify between the either and then use pretrained models to classify dog breeds.

## Solution Statement

The solution to the given problem statement would involve 2 classifications :

a) Classify between a dog and a human, if not either return error message
b) If the classification in (a) results in a dog, identify the breed of the dog.

The model should be able to consistently do both classification tasks irrespective of position of subject in image. Should exhibit translational, rotational invariance properties

It is assumed that the given image contains only a dog or human or other subject. The model is not expected to given with an input having, say, image of a dog with human. The performance of the model can directly be assessed by checking for accuracy, precision and recall.

## Benchmark Model

The benchmark model for this would be simply a probability predictor. So, if say there are 100 breeds of dogs to be identified then benchmark to beat would be getting >1% accuracy.

## Evaluation Metrics

The model will be evaluated on the basis accuracy of classification. I also plan to monitor precision and recall to ensure that our model is overfitting the data

## Project Design

*The project would consist of 3 different elements :*

*a)* Using pretrained model/ CNN Model to classify between a human and a dog.
*b)* Using pretrained model to classify specific dog species
*c)* Creating a webapp where user can provide an input image

From user's perspective, an image input would be provided. Then using AWS Lambda and AWS API, the image will be provided to our endpoint for inference. The endpoint would send the data to our model to do both classifications (human/dog and dog breed), then return results of classification tasks to endpoint. Which in turn would return the output to user via webapp