Machine Learning Engineer Nanodegree - Mohammed Ehsan Ur Rahman

Domain Background

A computer vision project which is the subset of deep learning techniques used and applied to solve traditional image processing tasks. The most used algorithms and architecture are neural networks and backpropagation. Recognising the breed of dogs is a challenging task in image classification sub branch of computer vision. There are hundreds of breeds in existence which are grouped into 10 distinct groups according to physical characteristics. Classification is a complex task as there are lots of classes and just subtle differences between each class. It is also computationally and memory wise expensive due to the number of features involved. Remembering all the breeds and distinguishing between similar breeds cannot be easily done by even humans. The current task involves two subparts one is image recognition, and the other is image classification which uses machine learning techniques like convolutional neural network.

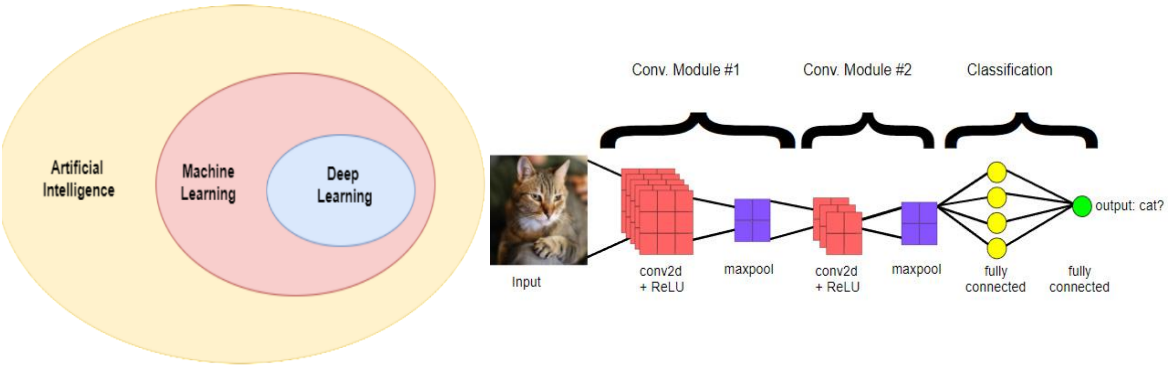

Figure 1 Domain of Deep Learning is a subset of Machine Learning which is a subset of Artificial Intelligence and Figure 2 A typical model of CNN

Problem Statement — a problem being investigated for which a solution will be defined

The aim of the project is to build a pipeline to process real-world, user-supplied images. The algorithm will identify an estimate of the dog's breed given an image. When the image is of a human, the algorithm will choose an estimate of a dog breed that resembles the human. If neither a dog or a human is detected,
then an error message is output. Therefore, the models in place should be capable of detecting a dog or human in an image, classify the dog to its breed and classify a dog breed that the human resembles.

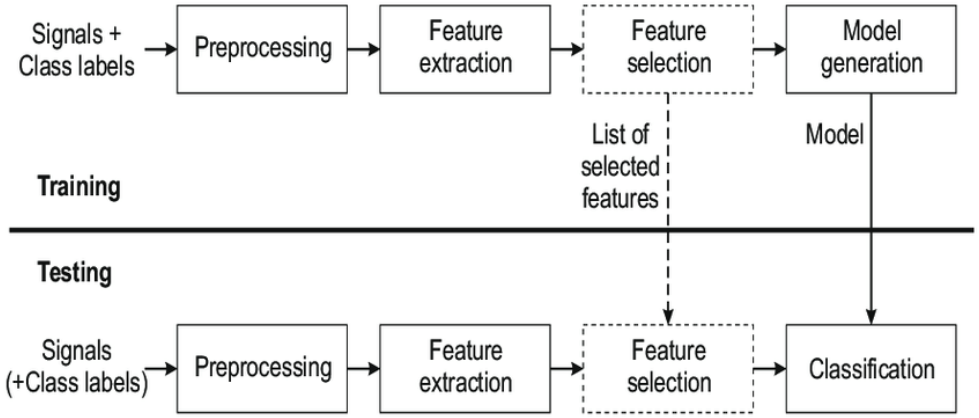

Figure 3 Workflow of training and testing

Datasets and Inputs — data or inputs being used for the problem

There is the dog dataset containing 133 breeds and each breed contains 8 images. There is also the human dataset[3] which contains images,_rst names and last names. There is a total of 13233 human images and 8351 dog images. The human dataset will be used to detect human faces in images using OpenCV's implementation of Haar feature-based cascade classifiers. The dog dataset is used to detect dogs in images using a pre-trained VGG-16 model. The datasets of human and dog images is justified for this project given the problem statement above.

The dogs' datasets are downloaded from the link dog dataset and placed it in the repo, at location path/dog-project/dogImages. The dogImages/ folder will be containing 133 folders each having images of respective breed's dogs, and each folder corresponding to a different dog breed.
The humans' datasets are downloaded the human dataset and placed it in the repo, at location path/dog-project/lfw.

Figure 4 Dog of breed: Affenpinscher

Figure 5 Human Face

Solution Statement — the solution proposed for the problem given;
1. Firstly a Haar cascade based built-in face detector is used in the project that is an OpenCV's implementation of Haar feature-based cascade classifier. It is used to detect human faces from the given image.
2. A pre-trained VGG-16 based model with trained weights on ImageNet, will be used to detect dogs. The pre-trained model should be checked that it returns indexes from 151 to 268 inclusive, that is, it must include categories from 'Chihuahua' to 'Mexican Hairless'.
3. A CNN model is then built from scratch to classify dog
breeds, that is, transfer learning cannot be used just yet. This model should surpass a test accuracy of 10% set by Udacity because the model is being built from scratch so classifying similar breeds can be a challenge however transfer learning will greatly improve this.
4. Finally, a transfer learning will be used with a ResNet50 model to significantly boost the accuracy of the CNN model. It should surpass the 60% test accuracy set by Udacity.

Benchmark Model — some simple or historical model or result to compare the defined solution to;
The pre-trained VGG-16 model l with trained weights on ImageNet, a large popular dataset for image classification, will be used as the benchmark model. ImageNet has over 10,000,000 URLs that link an image containing an object from one of 1,000 categories. The pre-trained model should be checked that it returns indexes from 151 to 268 inclusive, that is, it must include categories from 'Chihuahua' to 'Mexican Hairless' as stated in the solution statement.

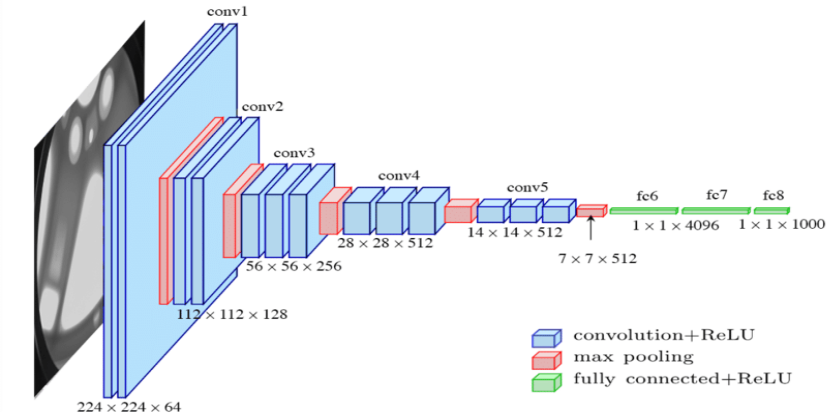

Figure 6 VGG-16 Architecture

Set of **Evaluation Metrics** — functional representations for how the solution can be measured;
Accuracy is the metric used to test both the benchmark model and
The my solution model.
Accuracy = TP+TN/(TP+TN+FP+FN)
True Positive(TP)= an outcome where the model correctly predicts the positive class
True Negative(TN)= an outcome where the model correctly predicts the negative class
False Positive(FP)= an outcome where the model incorrectly predicts the positive class &
False Negative(FN)= an outcome where the model incorrectly predicts the negative class.

An Outline of the **Project Design** — how the solution will be developed, and results obtained.
The outline of the project is given below:

1. The datasets for the human and dog images are imported and the total number of dog and human images is determined.
2. Haar feature-based cascade classifiers' OpenCV implementation is used to detect human faces from the datasets.
3. Pre-trained VGG-16 model is used to detect dog images from the datasets.
4. CNN model is created from scratch to classify the images. Using transformations of resizing and cropping the images to the size 224x224 pixels.
5. CNN model is created using transfer learning to classify the images and using the transformations of resizing and cropping the image.
6. Switch cases are developed that returns the predicted breed if a dog is detected. If a human is detected, it should return the resembling dog breed. If neither dog or human is detected, then error message is output.
7. The model is tested on images from the datasets to see the model is being used properly in the Application.

References:

1. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems* 25 (2012): 1097-1105.
2. LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." *nature* 521.7553 (2015): 436-444.
3. Liu, Jiongxin, et al. "Dog breed classification using part localization." *European conference on computer vision*. Springer, Berlin, Heidelberg, 2012.
4. He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
5. Cuimei, Li, et al. "Human face detection algorithm via Haar cascade classifier combined with three additional classifiers." *2017 13th IEEE International Conference on Electronic Measurement & Instruments (ICEMI)*. IEEE, 2017.
6. https://neurohive.io/en/popular-networks/vgg16/
7. https://iq.opengenus.org/resnet50-architecture/
8. https://medium.com/@erika.dauria/accuracy-recall-precision-80a5b6cbd28d#:~:text=Accuracy%20is%20an%20evaluation%20metric,True%20Positives%20and%20True%20Negatives.
9. Emami, Shervin, and Valentin Petrut Suciu. "Facial recognition using OpenCV." *Journal of Mobile, Embedded and Distributed Systems* 4.1 (2012): 38-43.