**Assignment Report**

***Image Recognition***

**Rohit Raj**

**12th May 2019**

**Submitted to : Exact space**

**Contents**

**1. Introduction**

1.1 Problem Statement . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. . . . . . . . . . 3

1.2 Data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

**2. Methodology**

2.1 convolution . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5

2.2 Apply the Relu (Rectifier Linear unit) function. . . . . . . . . . . . . . . . . . . . . .6

2.3 Pooling . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7

2.4 Flattening . . . . . . . . . . . . . . . . . . . . . . . . .. . . . . . . . . . . . . . . . . . . . . . . . . ..8

2.5 Full connection . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8

**3. Conclusion**

3.1 Model Evaluation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. .9

**References . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .** 9

**Chapter 1**

**Introduction**

* 1. **Problem Statement**

Objective:

You will be provided with two audio video clips in the form of mp4 files.

You are required to:

1. Extract image frames from each video and store them in folders “asset1” and “asset2”

2. Create a third folder “test”

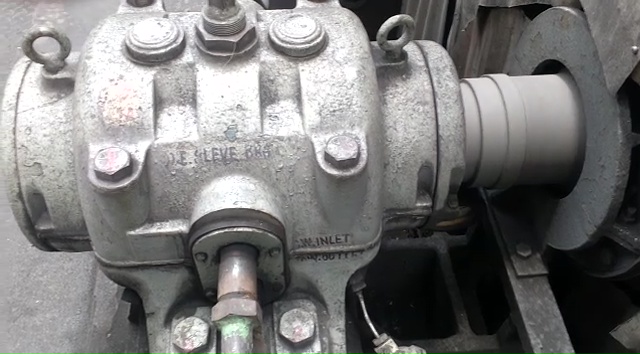
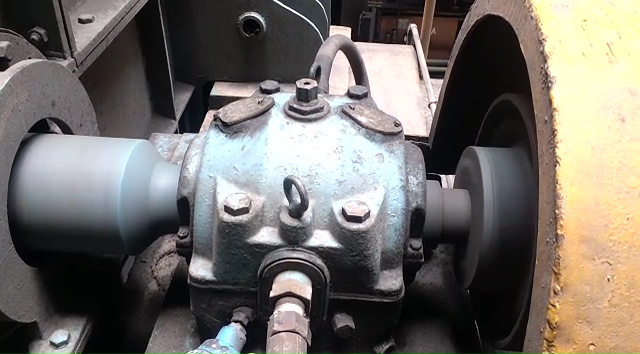
3. MOVE five images from “asset1” and asset2” to the “test” folder.

4. Using the images in “asset1” and “asset2” train an image recognition model that will be able to distinguish images between images in the two folders

5. Apply the model on the ten images stored in “test” folder

* 1. **Data:**

Please download data from <https://dev.exactspace.co/dt/videos.zip>

Asset 1:- Asset 2:-

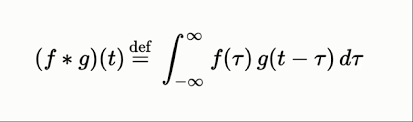
**Training set:** Asset 1 contains 352 images & Asset 2 contains 420 images

**Test** **set :** Both Asset 1 and Asset2 contains 5 images each.

**Chapter 2**

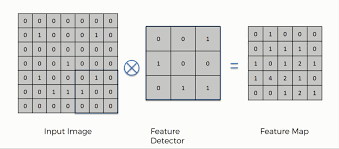
**Methodology**

**2.1 Convolution:** A convolution is a combined integration of two functions that shows how one function modifies the other,



There are 3 important items need to mention in this process:

* Input Image : - The input image is the image being detected.
* Feature Detector: - It is a matrix usually 3 cross 3 , (It could also be 7 cross 7 or 5 cross 5 ), also called (kernel or filter)
* Feature map: It is the image with extracted important features & with reduced size.



In above fig, Intuitively, the matrix representation of the input image is multiplied element wise with the feature detector to produce feature map. The aim of this step is to reduce the size of the image and make preprocessing faster and easier. Some of the features of the image are lost in the step. However, the main features of the image that are important in image detection are retained. These features are the ones that are unique to identifying the object.

**2.2 Apply the Relu (Rectifier Linear unit) function:**

In this step we apply the rectifier function to increase non-linearity in the CNN.

Images are made up of different objects that are not linear to each other. Without applying this function the image classification will be treated as a linear problem while it is actually a non-linear one.

Rectifier Linear unit is the most commonly used activation function in deep learning models. The function returns 0 if it receives any negative input but for any positive value x it returns that value back.

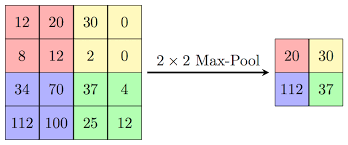
It can be written as –

f(x) = max(0,x)



**2.3 Pooling:** A pooling layer is another building block of a CNN. Its function is to progressively reduce the spatial size of the representation to reduce the amount of parameters and computation in the network. Pooling layer operates on each feature map independently.

The most common approach used in pooling is max pooling.

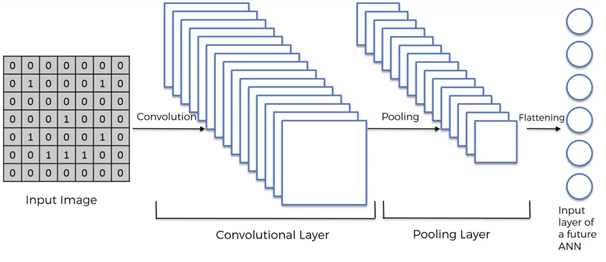


Max pooling works by placing a matrix of (2 cross 2) on the feature map and picking the largest value in that box. The (2 cross 2) matrix is moved from left to right through the entire feature map picking the largest value in each pass.

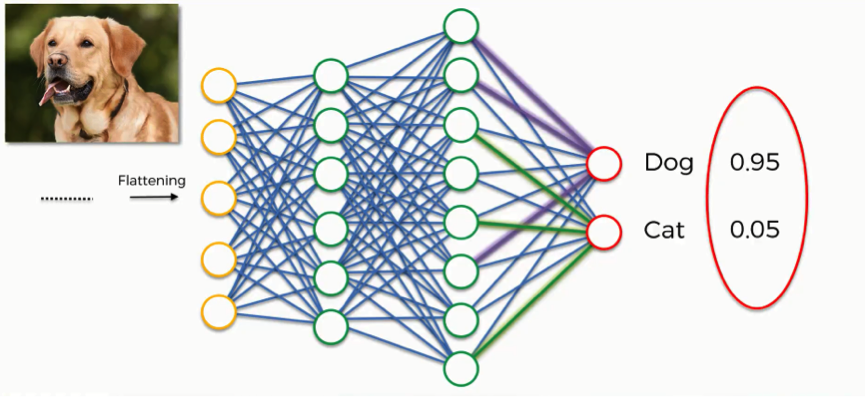
These values then form a new matrix called a pooled feature map. Max pooling works to preserve the main features while also reducing the size of image. This helps reduce over fitting.

Spatial Variance is a concept in pooling where the location of an image doesn’t affect the ability of the neural network to detect its specific features. Pooling enables the CNN to detect feature in various images irrespective of the difference in lightning in the pictures or different angles of the images.

**2.4 Flattening :** Once the pooled feature map is obtained, the next step is to flatten it. Flattening involves transforming the entire pooled feature map matrix into a single column which is then fed to the neural networks for processing.



**2.5 Full connection:** After flattening , the flattened feature map is passed through a neural network. This step is made up of the input layer, the fully connected layer (Hidden layer) and the output layer. The output layer is where we get the predicted classes. The information is passed through the network and the error of prediction is calculated. The error is then back propagated through the system to improve the prediction.



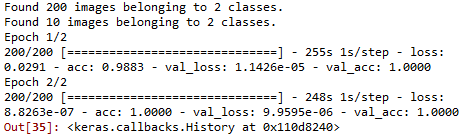
**Chapter 3**

**Conclusion**

**3.1 Model Evaluation**

**Training data accuracy = 100%**

**Test data accuracy = 100%**

****

There is no difference in the accuracy of Training and test data set. That means my CNN(Convolutional Neural Network) Model is not over fitting.

**Resources:**

[**https://www.udemy.com/deeplearning/**](https://www.udemy.com/deeplearning/)