# Image Recognition by Deep Learning

BRAC UNIVERSITY

Inspiring Excellence

by

Karishma Mohiuddin (13101137)

Amit Kishor Das (13301096)

Habiba Bint Obaid (13301026)

**Supervisor: Professor Md. Haider Ali**

**Co Supervisor: Dr. Jia Uddin**

*School of Engineering and Computer Science*

*Department of Computer Science & Engineering*

Date: 21 August, 2017

# DECLARATION

We, hereby declare that the thesis entitled “Image Recognition by Deep Learning” is based on image recognition under the supervision of Professor Dr. Md. Haider Ali and co supervision of Dr. Jia Uddin which is part of the degree of Bachelor of Science in Computer Science. All information has been presented in according with academic rules and ethical conduct and neither in whole or in part, has been previously submitted for any degree. Moreover, materials of work used here found by other researchers are fully sited and referenced.

Signature of Supervisor

**Professor Dr. Md. Haider Ali**

Signature of Co Supervisor

**Dr. Jia Uddin**

Signature of Author

**Karishma Mohiuddin**

Signature of Author

**Amit Kishor Das**

Signature of Author

**Habiba Bint Obaid**

# ACKNOWLEDGEMENT

Firstly, we would like to thank our thesis advisor Professor Dr. Md. Haider Ali of the Department of Computer Science & Engineering at BRAC University. Whenever we had faced any kind of problem or had a question about our research or writing, the door to his office was always open for us. He consistently allowed this paper to be our own work, but always guided us in the right direction. We are deeply grateful to him.

Moreover, we would like to extend our sincere gratitude to our Co supervisor Dr. Jia Uddin of the Department of Computer Science & Engineering at BRAC University for encouraging us, supporting throughout this thesis and giving us valuable suggestions.

Finally, we are grateful and we would like to express our heart-felt gratitude to the faculties, friends from BRAC University and family members for the influences and support. This accomplishment would not have been possible without the appreciation, guidance and help from them. Undoubtedly, it was a great inspiration and motivation in this thesis journey. Thank you.

# INDEX

# LIST OF FIGURES

# LIST OF TABLES

# ABSTRACT

Object recognition has become a crucial topic in the field of computer vision. Poor qualities of images unable bring out the desired object as per expectancy. Many models have proposed to recognize object from image. However, most of these approaches hardly achieve high accuracy and precision. It creates a major obstacle to get correctness of the research because of the lighting, illumination, image quality, noise, ethnicity and various angels of similar objects. Therefore, we have proposed a novel approach to detect any object by CNN method including HAAR Cascade classifier where we first detect the most prominent features from scene using Haar Feature Based Cascade Classifier that has been introduced by Paul Viola and Michael Jones. In the second phase, the classification has been used for Convolutional Neural Network to detect the object automatically with better accuracy and more efficiently. It can determine any object after proper training and dataset manipulation. Our proposed method for image recognition has achieved very good accuracy than our expectation.

Chapter 1

# INTRODUCTION

# 1. INTRODUCTION

It is important to detect specific kind of object from images nowadays. Our research will help out in various sectors like surveillance system, criminology, security and weaponry system. Similar type of object coming with various figures can easily be identified by human intelligence but it needs proper training to examine the object precisely and identify it perfectly by machine. For this, machine learning with deep learning approach is required with the help of Convolutional Neural Network [2]. The world is getting machine dependent in this modern era. As a result, object detection from images has become a major theme in the field of computer vision and image recognition fields. To detect object with good accuracy we have introduced deep learning method called Convolutional Neural Network along with HAAR Cascade classifier to detect the object with lesser errors[1][3]. Further, it will take less time and be more efficient than the previous works for object detection.

## 1.1 Motivation

As per previously mentioned "Deep Learning Approach" has enabled us to find out any object from picture by constructing layers of prominent features which are the most essential and important features to identify the object precisely[5]. Most of the pictures are taken nowadays are not in high quality or the image contains extra noise, blurry or lack of good lighting. This limitation hinders the machine to find object smoothly and even sometimes because of the quality of pictures it is very difficult to determine object with the human eyes. Our primary goal is to detect similar categorical objects from any type of image and determine that object based on the most important features that it has relied on.

## 1.2 Aims and Objectives

The primary aim of this thesis is to apply deep learning approach for image recognition with maximum accuracy. There has been a lot of work done in this field using various methods which all have their shortcomings. We have worked on Haar Cascade classifier for making classifier for true and false images [1] [9]. Besides, for image processing and recognition, deep learning can easily be applied with great success. We study different kind of deep neural networks algorithm and train the procedure with the art of Convolutional Neural Network [11]. For conducting our research we have collected raw data from internet manually and used it as dataset for the work flow. We have approached with few sequential steps to reach our aim. Though there are other procedures to frame the research work but we have approached with this manner to perform our research work as we have recognized it as an innovative way to do our work in the computer vision field. We have followed some steps to do our work.

Firstly, we have used Viola and Jones HAAR Cascade Classifier algorithm for separating false and true image to create a classifier for our research purpose.

Secondly, we have trained the dataset with CNN with 70% as trained set and 30% as test set to create CNN model.

Finally, we have tested our model with other images to recognize the pattern and detect the object with good accuracy as per our goal.

This thesis has specifically targeted on the issue of image recognition so that we may easily find desired object from any kind of classified image.

## 1.3 Thesis Outline

The thesis is ordered as follows:

- Chapter 1 is the discussion of proper prologue of the thesis which includes our inspiration for starting this thesis and goals and objectives for it.
- Chapter 2 has discussed about the literature review in where we have take related and reliable articles for our work which also discuss about theoretical approach
- Chapter 3 is the main theme of our work flow and work model that how our work has done.
- Chapter 4 represents the result and analysis after we have done our thesis and also discuss about the data flow how it works.
- Chapter 5 the conclusion part in which we discuss about the limitation, our future plan.

Chapter 2

# BACKGROUND STUDY AND RELATED WORK

# 2. BACKGROUND STUDY AND RELATED WORK

## 2.1 HAAR Cascade Classifier

### 2.1.1 Purpose of HAAR Cascade

As our system also relies on HAAR Cascade model, we have taken a basic idea from original author Viola and Jones [9] where they discussed about rapid boosted recognition of object using Cascade Classifier. In HAAR like features, some neighboring rectangular regions at specific location to add up the pixels intensities in each region of a fixed window. Therefore, after summing up the regions and calculating the differences between these sums, it is much easier to categorize the image [1]. In Cascade Classifier two types of data sets are needed, one is false image and another is true image set, with the proper training and execution, this classifier algorithm detect the image based on the region of interest [12].

**Figure 2.1**: Example of a “non-face” on the surface of mars [9]

## 2.2 Convolutional Neural Network

### 2.2.1 The purpose of CNN

By comparing with real time neuron from biological perception, a neural network is a structure of consistent artificial “neurons” which is supposed to exchange message among each other [5 ]. The connections are processed by numeric weights which enable to respond without errors when it is examined with certain image or pattern to recognize. The neurons are of multiple layers. Every layer consists of many neurons which are responsive to various combinations of input from the past layers [5].

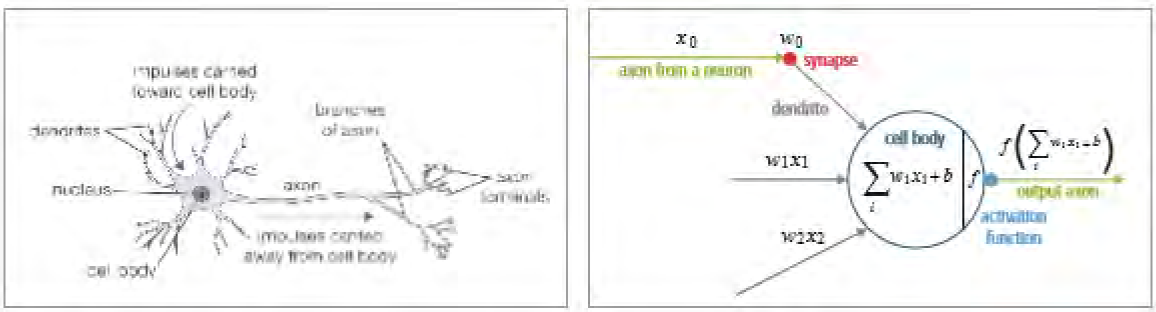


**Figure 2.2**: Illustration of a biological neuron (left) and its mathematical model (right) [5].

Training is measured by some label of trained dataset. By using general-purpose methods, training uses to iteratively find out the weights for transitional and very last feature of neurons. Figure 3 shows the training process at a block level [13].

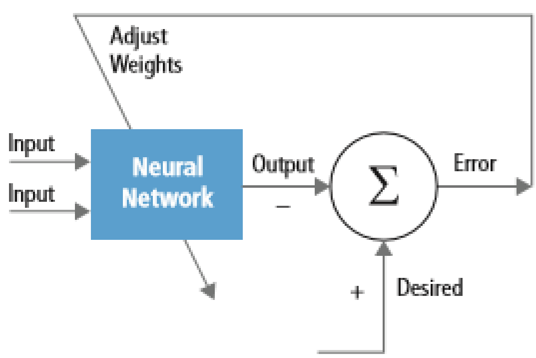


**Figure 2.3:** Training of neural network [5].

The proposed model has opened the door of new technique to recognize images. After conducting comparison with our proposition with others models, we have achieved a promising result. HAAR Cascade based classifier with CNN which has provided very good accuracy with fewer epochs. Authors in their research work [7], implemented CNN on training set of 4654 images with epoch of 600 and they had achieved more than 90% accuracy in total. As their dataset is large and had managed to perform many epochs, they have ended with very good accuracy. In article [14], the authors have used Deep Convolutional Neural Network to detect object with the data size of 1650 images and epochs of 25. They had achieved 60.74% of accuracy. On the other hand, our proposed model has achieved 88.2% of accuracy with only 5 epochs and data size is of 200 images. Certainly, it indicates that our model has more robustness and can achieve better accuracy with data expansion technique.

Chapter 3

# DESIGN APPROACH

# 3. Proposed Model and Workflow

In our proposed model, at first our initial task was inspired by Viola and Jones original work [9]. After using HAAR like cascade we have come up with classifier that has again used as training set for CNN in which 70% of data set is training set and 30% of data is used as test set [6]. This time, we train the model according to CNN convention and again make a classifier model which can accurately determine the difference between two unique objects.

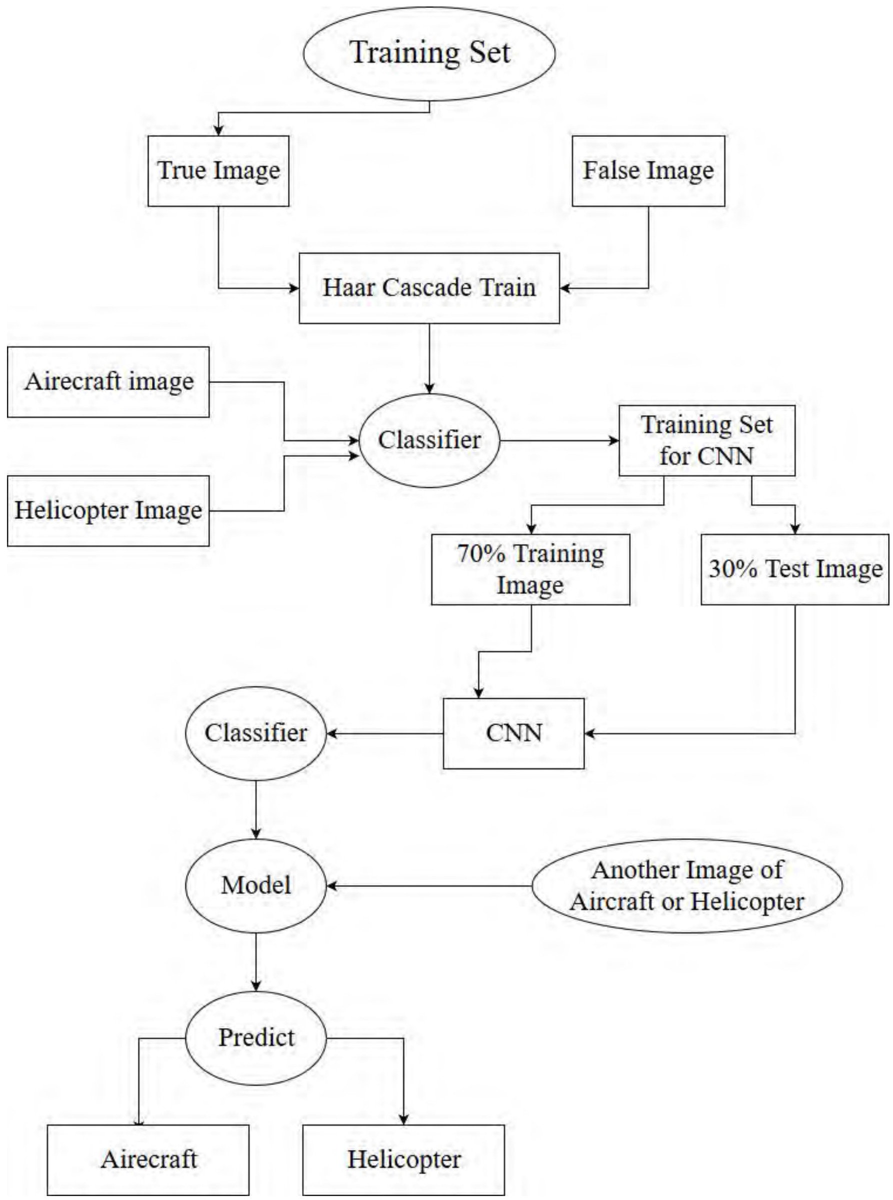


**Figure 3.1**: Proposed Model of Image Recognition

## 3.1 Creating First-Hand Classifier

### 3.1.1 HAAR Cascade Classifier

In HAAR-like feature, all the neighbor regions in a detection window have been taken into the account for a specific location and in each region resulting pixel intensities are summed up. After that, the difference between the resulting sums tends to categorize all the following sections of the image [12]. However, there are several necessary steps to train the HAAR cascade classifier. We took help from [10] while creating our own classifier. Steps are given below:

1. In order to train properly, we have taken images of helicopter and aircraft from online to get positive images and a greater or equal amount of negative images. Assuming, the positive images as $\rho$ and negative images as $\mu$. As we are taking the equal greater or equal amount of negative image in respect to the positive ones, we consider to have n numbers of negative and m numbers of positive images,

$$\sum_{i=0}^{n} \mu \geq \sum_{i=0}^{m} \rho \qquad (1)$$

2. It is necessary to mark or highlight the positive images using cropping tools. Otherwise, other elements of a scene would also get selected for HAAR features along with the objects we want to detect and detection rate would decrease. Here, background reduction techniques can be used for more accurate result. Each of the positive image $\rho$ has been cropped accordingly based on the requirement for minimizing the noise factor N.

$$MINIMIZE(N) \leftarrow \sum_{i=0}^{m} CROP(\rho)_{Requirement} (2)$$

3. Thirdly, we have to create an array of vectors by using the cropped positive images. Basically, the array of vector file resides in a vector file.

4. Finally, we have trained our classifier by using these negative and cropped positive images for the detection purpose.

$$ClassifierC \leftarrow TRAIN(\sum_{i=0}^{m} CROP(\rho)_{Requirement}, \sum_{i=0}^{n} \mu) \qquad (3)$$

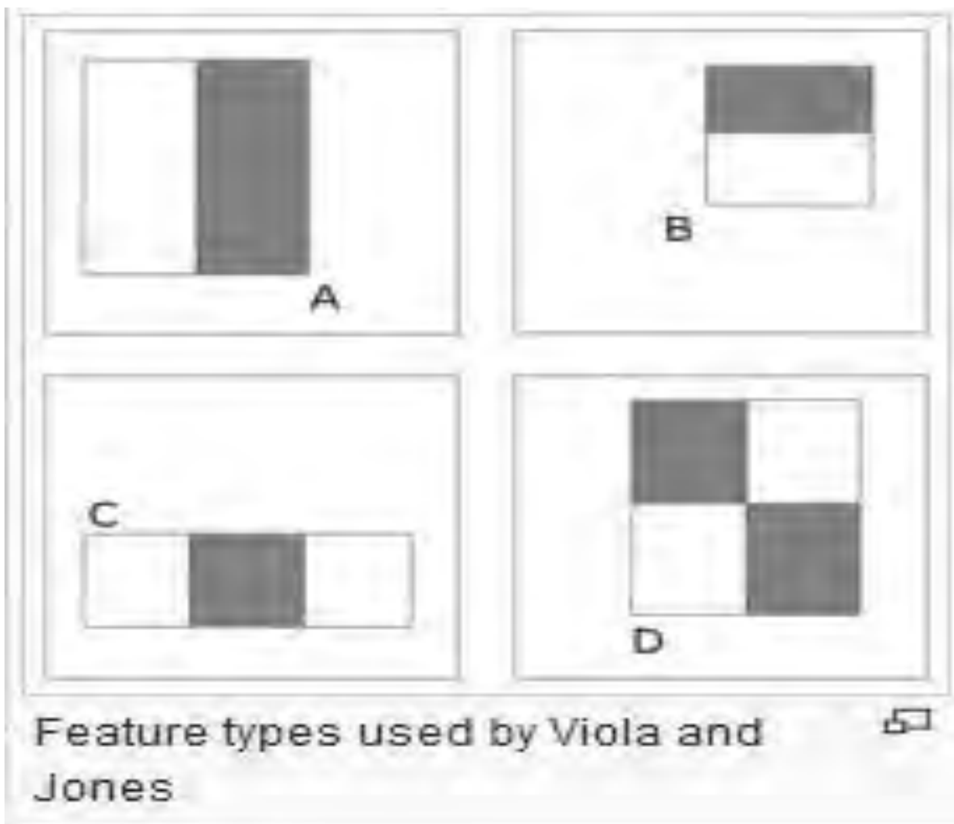


**Figure 3.1.1**: Haar Features [1]

## 3.2 Conversion in CNN

### 3.2.1 Creating Training Set

After we have created the cropped sized HAAR Cascade Classifier of every false and true image, we have proceed to develop further tasks by starting with airplane and helicopter images of about 250 units of each section. About 200 photos of helicopter and another 200 pictures of airplane have been taken by us for creating training set.


**Figure 3.2.1:** Training Set of Airplane


**Figure 3.2.2:** Training Set of Helicopter

### 3.2.2 Creating Test Set

To work our model of Convolutional Neural Network, a test set is also required for the procedure to be done. For this purpose we again have taken 50 images of airplane and helicopter. This test set will help to evaluate other images by comparing with it.

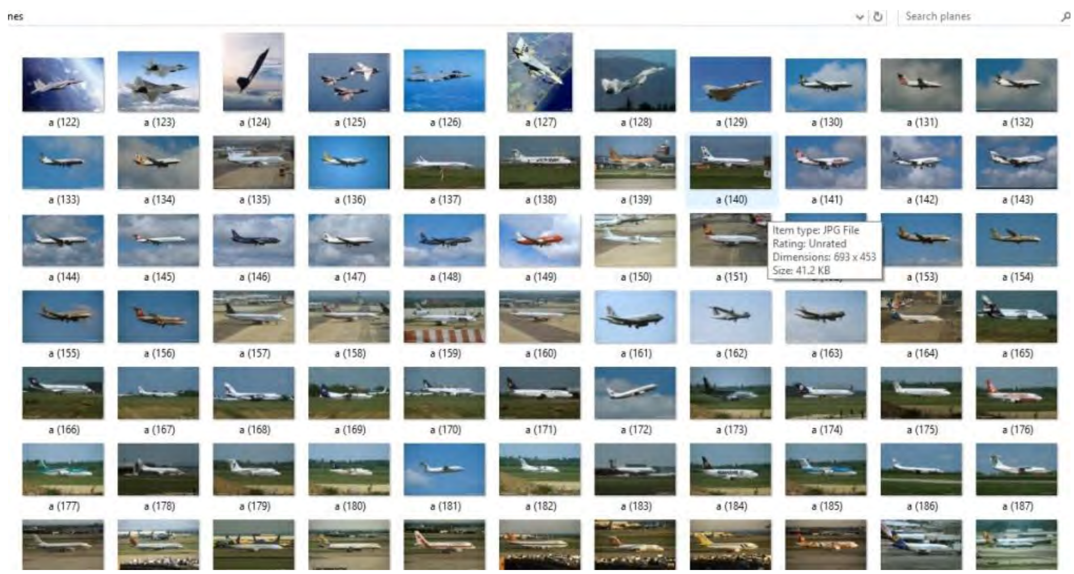

**Figure 3.3:** Test Set of Airplane

**Figure 3.4:** Test Set of Helicopter

### 3.2.3 Creating CNN Classifier

After 3 steps of layering process of Convolution we can transform low level features to high level features of each image then we have headed for pooling layer method [4]. The pooling layer decreases the resolution of the features and makes the features more robust against noise and alteration. After pooling layer, the images are shifted to flattening layer where all layers merge into single layer containing the most prominent features from 3X3 pixels of images. [7][8]

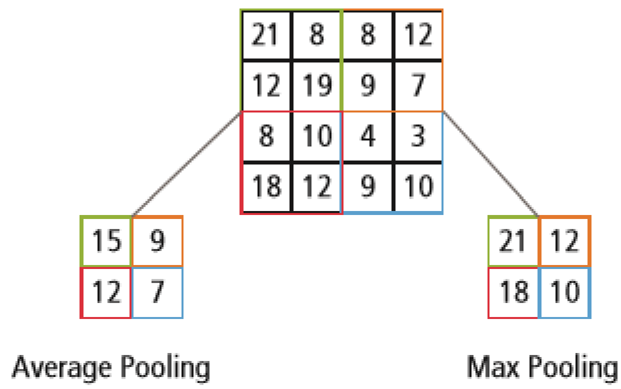


**Figure 3.5:** Representation of Pooling Layer [5]

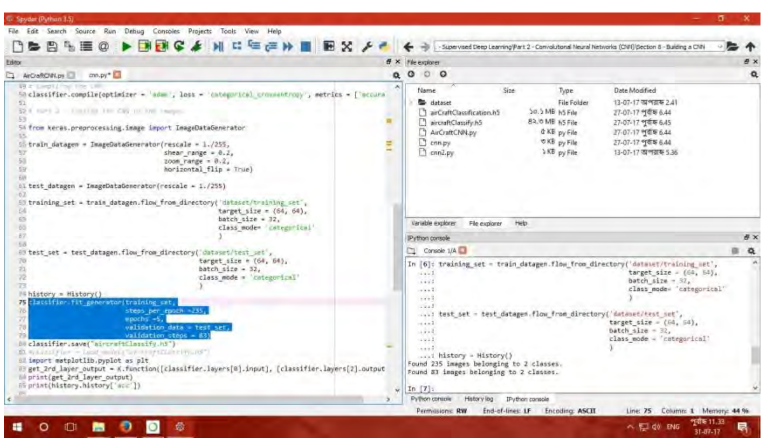

**Figure 3.6:** Training the Dataset with Deep Learning

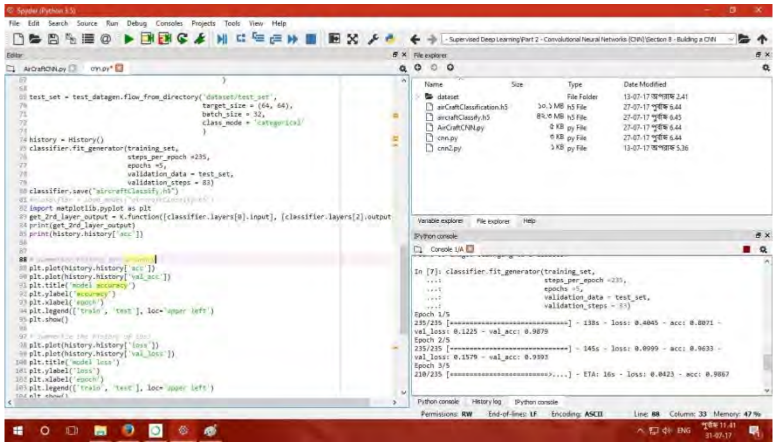

**Figure 3.7:** Creating the Model of CNN

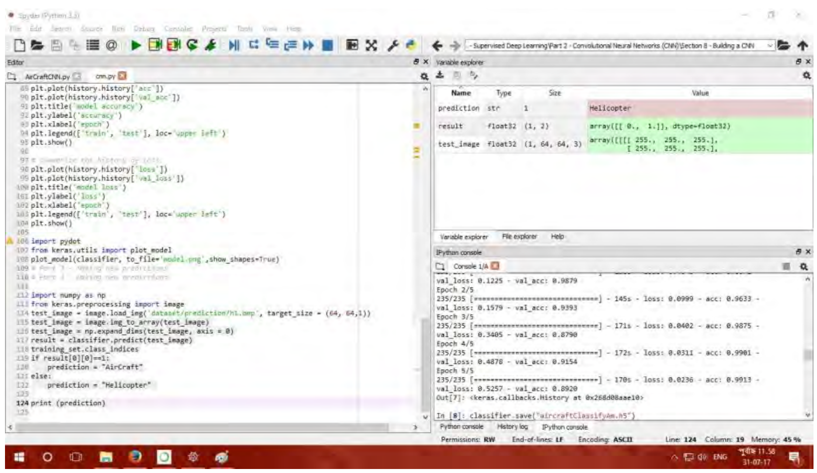

**Figure 3.8**: Predicting the Given Image

After creating the flattening layer, lastly, we proceed to last layer of CNN which is Fully-Connected layer. These layers summing up the weighting of the earlier layer of features that indicates the accurate mix of ingredients to verify a fixed target output result. In a fully connected layer, all the elements of all the features of the earlier layer are used for calculation of each element of each output feature [5].

Our working Model has showed the procedure of our work flow to detect the image.

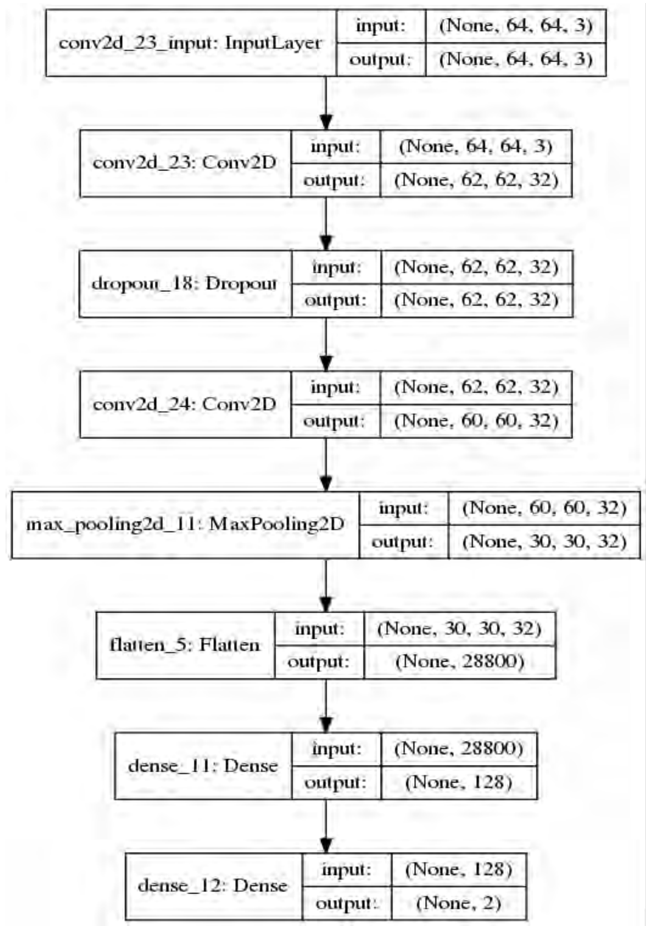


**Figure 3.9**: Model of Our Working Procedure

After creating the CNN classifier we can take it as our model set to test any other helicopter or airplane photos. This Model is ready to predict any kind of picture of two objects airplane and helicopter with good accuracy.

Chapter 4

# RESULT AND DATA ANALYSIS

# 4. RESULTS AND DATA ANALYSIS

In this work we have collected images by ourselves as primary raw data and also have used the source code of author [8] which is later modified by the requirement of our work. Our model has achieved 88.2% accuracy to recognize object such as helicopter and airplane.

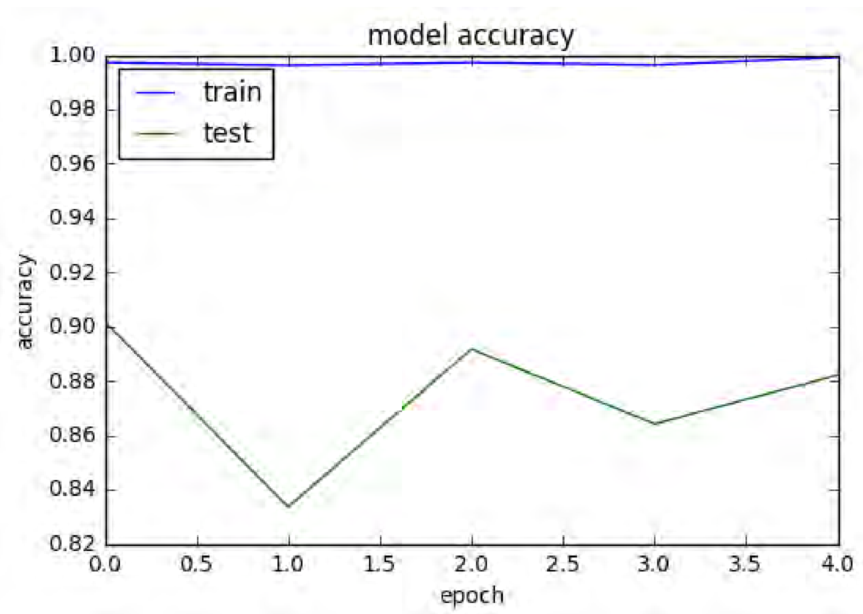


**Figure 4.1**: Graph of our Model's Accuracy

We can reckon that, while we have been training our test dataset, we could see the fluctuation of the total accuracy of our model to recognize the precise object. At the end point when our procedure has completed, the accuracy gradually build up and ended with 88% accuracy which is very good result in image recognition field.

We could also determine the net loss of our work to predict the actual accuracy and fineness of our model. At first, the percentage of loss good while we are training our model but later it has started decreasing steadily and ended at almost 77% which indicates that loss is less with much very good consistency than the expected prediction.

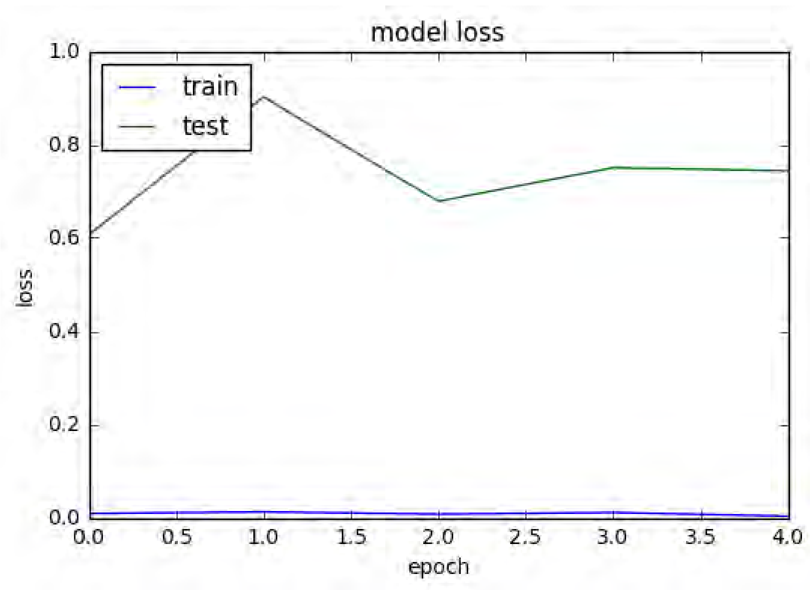


**Figure 4.2**: Graph of our Model's Loss

This certainly indicates that our model functions with good precision and less errors.

Here Accuracy,

$$Accuracy = \frac{TotalRecognized}{TotalInput}$$

| Round | Number of Image | Detection | | Accuracy of True Detection | Average Accuracy |
|---|---|---|---|---|---|
| | | True | False | | |
| 1 | 60 | 53 | 7 | 89% | 88% |
| 2 | 80 | 68 | 12 | 86% | |
| 3 | 50 | 42 | 8 | 85% | |
| 4 | 70 | 61 | 9 | 88% | |
| 5 | 100 | 92 | 8 | 92% | |

**Table 4.1**: Accuracy of True detection

We can take some images and then detect by HAAR Cascade classifier and after that when we implement CNN model on the test set, we observe a good accuracy with average more than 80% in every possible test set that we have created. We can determine the true detection with very good precision and then again by averaging all the accuracy we can certainly determine about 88% exactness of our model work.

Chapter 5

# CONCLUSION AND FUTURE PLAN

# 5. CONCLUSION AND FUTURE PLAN

In our paper, we have been worked on image recognition by deep learning with the help of HAAR Cascade classifier of main author Viola and Jones[9] and also as a part of deep learning, Convolutional Neural Network have been applied by us into it [11][13]. Though we have achieved much good accuracy with very good result still there are somewhat limitations that we have put aside for our future work and research. Again, it is a procedure to detect binary object whereas we could work it out for detecting multiple objects from the same images. Further, if we test the model with blurry or distortion picture then it cannot determine the targeted object of that specific picture. This drawback could also lead us to our future work to make our model more robust and more significant to recognize precise objects from the image. These issues will be looked forward to solve in proper research.

# References:


[1]Cascade Classification¶. (n.d.). Retrieved August 03, 2017, from http://docs.opencv.org/2.4/modules/objdetect/doc/cascade_classification.html

[2] Deshpande, A. (n.d.). The 9 Deep Learning Papers You Need To Know About (Understanding CNNs Part3. Retrieved August 03, 2017, from https://adeshpande3.github.io/adeshpande3.github.io/The-9-Deep-Learning-Papers-You-Need-To-Know-About.html

[3] Geitgey, A. (2016, June 13). Machine Learning is Fun! Part 3: Deep Learning and Convolutional Neural Networks. Retrieved August 03, 2017, from https://medium.com/@ageitgey/machine-learning-is-fun-part-3-deep-learning-and-convolutional-neural-networks-f40359318721

[4] Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., ...& Wang, G. (2015). Recent advances in convolutional neural networks. *arXiv preprint arXiv:1512.07108*.

[5] Hijazi, S., Kumar, R., & Rowen, C. (2015). *Using convolutional neural networks for image recognition*. Tech. Rep., 2015.[Online]. Available: http://ip. cadence. com/uploads/901/cnn-wp-pdf.

[6] LeCun, Y., &Bengio, Y. (1995). Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, *3361*(10), 1995.

[7] Lu, Y. (2016, December 03). Food Image Recognition by Using Convolutional Neural Networks (CNNs). Retrieved August 03, 2017, from https://arxiv.org/abs/1612.00983

[8] Object Recognition with Convolutional Neural Networks in the Keras Deep Learning Library. (2017, March 30). Retrieved August 03, 2017, from http://machinelearningmastery.com/object-recognition-convolutional-neural-networks-keras-deep-learning-library/

[9] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” Comput. Vis. Pattern Recognit., vol. 1, pp. I–511–I–518, 2001.

[10] Rezaei, M. (2015). Creating a Cascade of Haar-Like Classifiers: Step by Step.

[11] Simard, P., Steinkraus, D., & Platt, J. (n.d.). Best practices for convolutional neural networks applied to visual document analysis. *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings.* doi:10.1109/icdar.2003.1227801

[12] Soo, S. (2014). Object detection using Haar-cascade Classifier. *Institute of Computer Science, University of Tartu*.

[13] Wu, R., Yan, S., Shan, Y., Dang, Q., & Sun, G. (2015, July 06). Deep Image: Scaling up Image Recognition. Retrieved August 03, 2017, from https://arxiv.org/abs/1501.02876

[14] Zhang XJ, Lu YF, Zhang SH. Multi-task learning for food identification and analysis with deep convolutional neural networks. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 31(3): 489–500 May 2016. DOI 10.1007/s11390-016-1642-6