IMAGE ANNOTATION FRAMEWORK

A report submitted to

M S RAMAIAH INSTITUTE OF TECHNOLOGY

Bengaluru

IS811 SENIOR PROJECT

as partial fulfilment of the requirement for

Bachelor of Engineering (B.E) in Information Science and Engineering

by

Avinash R (USN- 1MS13IS026)

Apoorva G (USN- 1MS12IS021)

Gurushankar S G (USN- 1MS14IS403)

under the guidance of

Mr. Prashanth Kambli

DEPARTMENT OF INFORMATION SCIENCE AND ENGINEERING

M S RAMAIAH INSTITUTE OF TECHNOLOGY

May 2017

Department of Information Science and Engineering
M S Ramaiah Institute of Technology
Bengaluru - 54

CERTIFICATE

This is to certify that Avinash R (USN- 1MS13IS026) , Apoorva G (USN- 1MS12IS021) AND Gurushankar S G (USN- 1MS14IS403) who were working for their **IS811 SENIOR PROJECT** under my guidance, have completed the work as per my satisfaction with the topic **IMAGE ANNOTATION FRAMEWORK**. To the best of my understanding the work to be submitted in dissertation does not contain any work, which has been previously carried out by others and submitted by the candidates for themselves for the award of any degree anywhere.

(Guide)
Mr. Prashanth Kambli
Assistant Professor

(Head of the Department)
Dr. Vijay Kumar B P
Professor & Head, Dept. of ISE

(Examiner 1)

Name

Signature

(Examiner 2)

Department of Information Science and Engineering
M S Ramaiah Institute of Technology
Bengaluru - 54

DECLARATION

We hereby declare that the entire work embodied in this **IS811 SENIOR PROJECT** report has been carried out by us at M S Ramaiah Institute of Technology under the supervision of Mr. Prashanth Kambli. This Project report has not been submitted in part or full for the award of any diploma or degree of this or any other University.

Avinash R(USN- 1MS13IS026)
Apoorva G (USN- 1MS12IS021)
Gurushankar S G (USN- 1MS14IS403)

Acknowledgements

This project would not have been possible without the support of many people, each of whom, in their own capacities has helped us in carrying out our work. We would like to take this opportunity to thank them all. First and foremost we extend our hand to thank our
HOD, Dr. Vijay Kumar B.P, Professor and Head, Department of Information Science Engineering, for permitting us to take up this project. We also wish to express our gratitude to our project guide, Mr. Prashanth kambli who was abundantly helpful and offered invaluable assistance, exemplary guidance, monitoring and constant encouragement.

We would also like to thank all teaching and non-teaching staff of our department for their kind cooperation. Although there are many who remain unacknowledged in this humble note of gratitude, there are none who remain unappreciated.

Abstract

In the recent years Image Annotation has picked up a significance in the process of searching, retrieving and applying labels for images.The project involves in identifying and recognizing the objects within the given image using predefined neural network learning algorithms and tools. The model also includes recognition of complex images using partially annotated datasets for improved annotation process. The model accomplishes two major tasks: Initially, to explain with labels and later to segment the picture with the comparable area, such that the expectation of the overall model will be useful in computer vision.

The proposed model is carried out using MatLab and ImageNet dataset. The performance of the system demonstrates the presence of correctness within the inputted image using labels (tags) and partitioning pictures from different classes portraying to various scenes.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

With the development of the Internet and digital imaging devices many large image collections are being created. Popular online photo-sharing sites like Flickr contain hundreds of millions of diverse pictures. Many organizations, e.g. libraries, hospitals, governments and commerce have also been creating their large image databases by scanning paintings, manuscripts, prints and drawings. Searching and finding large numbers of images from a database is a challenging problem. Search engine they do not really capture the semantics or meaning of images well. For image retrieval systems based on text queries, the key problem is how to get the metadata such as captions, titles or transcriptions. Manual annotation is not practical for large volumes of image sets. Commercial image search engines for the Internet, e.g. Google image and Yahoo image, use the text surrounding each image as its description. However, these search engines entirely ignore the visual content of the images and the surrounding text doesnt always relate to the visual content of an image. The consequence is that the returned images may be entirely unrelated to users needs.

Vision is the richest sense that a human being has which computer does not have and will consist of a tedious process to achieve the same for a computer. Object recognition and classification play a major role in this field.So we need a framework which is used for detecting objects from the image and annotate them with the proper tags, which will be used for problems stated above.

1.2 Scope

Image annotation is a complex job of detecting objects and classifying each objects in a given image. Even though the process is extremely useful in some cases, the complexity of the process limits many novice developers from using image annotation and object detection in their projects. So we are developing a user friendly framework that will do simplest of the image annotation tasks and help novice developers in their projects that might need object detection and classification.

1.3 Problem Statement

Annotating images on a small scale for personal usage, as described in the first use case above, can be relatively simple, as exemplified by popular keyword-based tagging systems such as Flickr. Unfortunately, for more ambitious annotation tasks, the situation quickly becomes less simple. Larger scale industrial strength image annotation is notoriously complex. Trade offs along several dimensions make professional multimedia annotations difficult:

- Manual annotation. It takes lots of time and people skill to annotate millions of images that are found in the internet.
- A large number of classes to which a single object can belong.In an image there might be more than one number of objects present in it, deciding to which class it will belong will be an ambiguous task.
- Classes are not usually combination of one or more classes. Class Labels are not usually having the property of the other classes, it will be distinct from one class to another class.
- Manual versus automatic annotation and the "Semantic Gap". In general, manual annotation can provide image descriptions at the right level of abstraction. It is, however, time consuming and thus expensive. In addition, it proves to be highly subjective: different human annotators tend to "see" different things in the same image. On the other hand, annotation based

on automatic feature extraction is relatively fast and cheap, and can be more systematic. It tends to result, however, in image descriptions that are too low level for many applications. The difference between the low level feature descriptions provided by image analysis tools and the high level content descriptions required by the applications is often referred to, in the literature, as the Semantic Gap.

- Weak Labeling and Multiple Instance Learning. The existing images on the internet lacks information about the documents that is metadata and also there will be multiple instances of an image that will cause the ambiguity.
- Noisy Images. Some images on internet are blurry also low dimension and low quality images are very difficult to process the image for automatic annotation.

1.4 Objectives

1. To provide an user friendly graphical interface to input the image.
2. User should be able to upload image
3. System should be able to pre-process the given input.
4. System should detect objects present in the image.
5. System should be able to Annotate the image with with appropriate tag.

1.5 Applications

Following are the application of Object Recognition and Annotation:

1. Android Eyes Object Recognition: Its an advanced object recognition application. If we take picture of an object then machine eye can tell what it is.
2. Automated vehicle parking systems: it is designed to reduce volume or area required for parking cars.

3. Optical Character identification: it is the conversion of image typed,handwritten in to encoded text.

4. Content-Based Image Indexing: Is searching for digital images in huge database. Search analyses the contents of the images.

5. Image watermarking: It is the method of smacking the digital information. Its used to find ownership.

6. Global robot localization: It is constructing and updating a map of environment which is unknown also simultaneously keeping track of the location of the agent.

7. Face detection: Used to recognize human faces. Using this process person locate and attend to faces in a scene.

8. Video Stabilization: helps in cancelling our moves while we are recording a movie.

9. Manufacturing Quality Control: It ensures that our products are safe, pure and effective. And products are released only after though analysis as per specification.

1.6 Challenges

Image Annotation is extremely useful in many areas of computer vision such as MultiMedia understanding, machine learning, image processing and analysis, and also in the field of querying and retrieving the proper information. When we try to analyze an image and annotate them with the caption, it's usually done with feature vector calculation and will be required to train a lot of training model to predict words/tag/words using machine learning technique to predict. Annotation of new image will be possible only after training and learning of the model. The task of scene understanding and object recognition for semantic prediction is challenging work. In Image Annotation Process we deal with recognizing multiple object.

1.7 Existing System

There are so many Opencv Version 3.0 for programming, which has all the kinds built in algorithms and libraries for image feature extraction, segmentation and also further processing of the images. OpenCV is built to provide a common infrastructure for computer vision applications and to fasten Process of object detection by using the machine learning methodology. The OpenCV has more than 2500 algorithms which is optimized for Computer Vision Problems, which includes a comprehensive set of both classic and state-of-the-art computer vision and machine learning algorithms. These openCV algorithms can be used to recognise and detect car number plate, identify objects in image, differentiate human actions in videos, track camera and other movements, track objects that are moving, extract 3D models from an image and identify the object, stitch images together to produce a high quality image of an total scene, find similar images from an image dataset and match them, follow eye movements, recognize scenery and establish markers to overlay it with augmented reality, etc.

1.8 Proposed Model

Here we have used MatLab Computer Vision toolbox which is efficient and faster than OpenCV. Objective is, we are trying to extract feature from an image and annotate them with appropriate tags for the object present in them. We extracted features and classify the image using AlexNet (A pre-trained convolutional Neural Network) which uses SIFT(Scale-invariant feature transform) which is advanced method than ORB and SURF, which helps in reducing noise. And also it uses dropout technique to selectively ignore single neurons during training avoiding the averaging effects of average pooling. Benefits of proposed work are:

- Firstly, language selection is very important, MatLab is very efficient in terms of space and time.
- Secondly, we used AlexNet which is one of top most ImageNet Classification in the field of neural Network.

1.9 Organization of Report

Chapter 1 serves as an introduction to the project. It informs of the reasons as to why the project was chosen, what the scope of the project is, what are the outputs expected from the project, and how the implementation of the project was planned to be performed.

In Chapter 2, several papers that undertook work similar to this project are discussed in order to strengthen the project by overcoming the shortcomings, if any, present in the implementations mentioned in those papers.

Chapter 3 discusses the design of the system that would be necessary to implement the project. The descriptions as to how the hardware, software, and the storage must be designed are provided. It also deals with the functional and non-functional requirements that the project would be expected to meet as well as the form of inputs and outputs the project would be able to deal with.

Chapter 4 describes how the actual implementation of the project took place and how it fulfills the expectations set in the design phase. It describes the algorithm used to obtain the required results.

Chapter 5 examines the output of the project by comparing the result obtained against the theoretical solutions to ascertain whether the system is behaving as expected.

Chapter 6 concludes the report, and informs as to any limitations of the project and how these could be dealt with. It discusses on the feasibility of an extension of the project and what functionalities could be added to the system in the future.

Chapter 2

Literature Review

Literature survey helps us to make detailed study of the background of our project. This helps in finding the flow of the project and also solving the problem. Different research has been done for the scheduling. This section helps to find references to discuss about the topics related to our project. So following research will give the background of the project and also flaws which helps to find the solution.

2.1 Image Annotation framework survey

To solve the problem of captioning the image automatically. In this, a training model is developed with captioned images, and try to predict and caption the image correlation with objects and caption the image with appropriate tags. Here researchers try to find the relationship with the image feature extracted and the keywords, in order to find the appropriate keyword for the new image. The authors done the performance measure with the alternate design on large set of data and the proposed model achieves up to a 45percent relative captioning accuracy over the other design model [1].

Tradition of of manual image annotation for retrieving images and indexing the images for collection is very expensive in terms of labor and due to the millions of images over the internet it's has become a tedious job to annotate every single image on the internet. Hence, author proposes automatic way to annotate photograph for retrieval of images based on a set of pre trained image

data set. Author assumes that regions in a multimedia can be segmented into frames and detect the blobs of vocabulary. Blobs are image features, generated from process called clustering. A model will be trained with a set of images with captions, the resulting model will be given a new image to process the model will show the probabilistic set of words for a given set of blob. The resulting words can be used for automatically caption a images with its object and helps retrieve the more accurate image when for a query.This model concludes that it's six times good as any other model based on the word blob co occurrence model and twice good as the machine translation model[2].

System that rely on low quality image are limited to semantic gap problem, which will lead to mismatching between notion of similar and the one which is adopted by a system. The solution is to semi automatically annotate and assign meaningful multimedia, so that enables the concept based retrieval. So in this research tries to graph based link analysis technique in the development of semi automated image annotation system. The system will predict the tag based on the variable number, representing on related terms, and abstract terminology[3].

The multimedia classification and annotation are current playing important role in the research in the field of computer vision, but it's not usually considered together. Annotation or captioning provides evidence for the classification by means of class label or tags for annotation. For example, an image of class road is annotated with words Streetlight, car, and signal than words frame, books, and skate. In this research paper, author develops a probabilistic model for together describing and classifying the image with a annotations and tags. This model describes class label as general description of the image and tags as local description of the image. its underlying probabilistic assumptions naturally integrate these two sources of information[4].

Usually the annotation are taken from an annotation vocabulary of few hundred class labels. Because of large class labels, there is less chance of that a images corresponding to the appropriate tags and due to the lack of manual annotation, there a millions of images that are in the internet are poorly labeled.so the researcher proposes 2PKNN, which uses 2 steps classic K-nearest neighbour algorithm, that solves these issues in the image annotation work. In the first step of 2PKNN uses image to label is done, in the second step it uses image to image,

hence this to process uses nearest-neighbour based methods greatly depends on how features are compared. Researcher also propose a metric learning framework over 2PKNN that learns weights for multiple features as well as distances together. 2K PNN performs better to the current state of the art on the image captioning dataset and shows significant improvement over the other design[5].

2.1.1 Comparison on Image annotation work

SNo	Model Name	Author	Segmentation Type	Object Recognition	Feature Extraction	Annotation Type
1	Co-occurrence Model	Mori et al	Each image in training set is divided into non overlapped parts.	Statistical learning of image to word relationship.	Clusters are made by vector quantization technique.	Probability for each word in each centroid is estimated statistically and word is assigned to that part of image.
2	Translation Model	Duygulu	Region Based Segmentation.	K-means Vector quantization.	Blobs extraction and map to words	Word to blob is used to annotate.
3	Tag Propagation	Matthieu Guillaumin	NA	Weighted combination of the tag absence/presence among neighbors.	Based on distance weight technique.	Prediction based tagging and annotation.

4	Group Sparsity Model	S. Zhang	No Segmentation, sparse feature	Word to word matching is done.	sparse feature extraction technique	Annotation using group sparsely.
5	Maximum Entropy	J.Jeon	NA	Rectangular regions of image, clusters using k-means algorithm.	uniform distribution.	Annotation is based on maximum atrophy.
6	Coherent Language Model	R.Jin, J.Liu	NA	Word to word is used and word is related or mapped to image.	Word to word relation.	Active learning is used generally in iterative manner.

Table 2.1: Comparison on Image Annotation Work

2.2 Object Detection and Recognition

Object detection and recognition are the first step in captioning the image, basically we find to find what all objects are present in the image and then we proceed with recognising the objects that are detected in the image. Here i list some basic MatLab functionality that will be used in the process of image annotation.

2.2.1 Object Detection in a Cluttered Scene Using Point Feature Matching

This algorithm works on detecting image based on finding point correspondences between two objects that are reference image and target image. It detect objects image even though the plane change or scale change. It Robust to little amount of out of plane rotation and conclusion. This method of object detection works

better on the images which exhibits non-repeating texture patterns,which give rise to unique feature matches.

2.2.2 Object Detection Using Deep Learning

This algorithm tells about how train object detector for detecting objects using R-CNN. R-CNN, basically it's a object detection tool box which uses technique called convolutional neural network (CNN) to classify the image regions within an image. This algorithm instead classifying every single pixel using a sliding window, the R-CNN method concentrates on those regions which has more percentage to contain an object. This will help in reducing the computation work and also time and space complexity. In transfer learning, a network trained on a large set data, such as ImageNet is used as the beginning point to solve classification or detection process. Using This approach that the pre trained network that already has the rich set of image features that are applicable to a large image dataset. This learning is transferable to the new task by fine-tuning the network.A network will be fine tunes by making adjustments to the heights such as the feature representations learned from the original process and adjust slightly to support the new process.

2.3 Convolutional neural network (CNN)

.

Its Basically, a type of feed forward artificial neural network in which the connectivity pattern between its neurons is inspired by the organization of the animal visual cortex. Individual cortical neurons respond to stimuli in a restricted region of space known as the receptive field. The receptive fields of different neurons partially overlap such that they tile the visual field. The response of an individual neuron to stimuli within its receptive field can be approximated mathematically by a convolution operation. Convolutional networks were inspired by biological processes and are variations of multilayer perceptrons designed to use minimal amounts of preprocessing. Convolutional neural network applications in image

and video recognition, natural language processing and also in recommender system(shopping cart).

2.3.1 Image Retrieval Based on Convolutional Neural Network

Retrieving similar image using image retrieval, and the effect of retrieving the image is depended on the image selection and feature extraction to some extent. But the author says based on the deep learning algorithm which has self learning ability of convolutional neural network to extract more helpful to next level semantic feature retrieval using CNN, and also to use as a distance metric function similar image. By surveying this paper we realize that convolution neural network model to extract the high level semantics of the image, and also image retrieval by analyzing the structure of the network Deep convolutional neural network firstly, the image is gradually learning and abstract, each layer can be generated to describe the image content of the underlying feature of the image[19]. Convolution neural network also advances in 3 Dimension Sensing technology making it easily to record the color and depth of the image which indeed helps in the object recognition in 3D modality.A model based on a combination of convolutional and recursive neural networks for learning features and classifying RGB-D images is developed. The CNN layer learns low-level translationally invariant features which are then given as inputs to multiple, fixed-tree RNNs in order to compose higher order features[15]. Deep hierarchical architectures accomplish the best published results on benchmarks for object recognition and with error rates of 2.53%, 19.51%, 0.35%, respectively. Deep nets trained by simple back propagation perform much better than more shallow ones. Learning is surprisingly rapid[17].

2.3.2 ImageNet DataSet Classification with Deep Convolutional Neural Networks

Deep convolutional neural network to classify the 1.2 million high quality images in the ImageNet LSVRC-2010 contest into the 1000 different classes/tags. On the test data, we obtained top-1 and top-5 error rate of 37.4 and 17.2 which is much better than the previous state of the art. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax. To make training faster, we used non-saturating neurons and a very efficient GPU implementation of the convolution operation. To reduce overfitting in the fully-connected layers we employed a recently advanced regularization method called dropout that proved to be very effective. We also entered a variant of this model in the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry.

We try investigate the effect of the convolutional network depth on its exactness in the large scale image recognition setting. Thorough evaluation of networks of increasing depth using an architecture with very small (3 3) convolution filters, which shows that a significant improvement on the prior-art configurations can be achieved by pushing the depth to 1619 weight layers[12].

We evaluate whether features extracted from the activation of a deep convolutional network trained in a fully supervised fashion on a large, fixed set of object recognition tasks can be re- purposed to novel generic tasks. Our generic tasks may differ remarkably from the originally trained tasks and there may be inadequate labeled or unlabeled data to conventionally train or adapt a deep architecture to the new job. We explore and visualize the semantic clustering of deep convolutional features with respect to a variety of such tasks, including scene recognition, domain adaptation, and fine-grained recognition challenges.we analyze the use of deep features applied in a semi-supervised multi-task framework[14].

How Convolutional Neural Networks, trained to know objects primarily in photos, perform when applied to more abstract representations of the same objects. Main goal is to better understand the generalization abilities of these networks and their learned inner representations. So both GoogLeNet and AlexNet networks are largely unable to recognize abstract sketches that are easily recognizable by humans. Here show that the measured efficacy vary considerably across different classes.The work presented here contributes to the understanding of the applicability of CNN in domains that are different but related to that of the training set [18].

Chapter 3

System Analysis and Design

3.1 System Requirement Specification

System Requirement Specification (SRS) is a central report, which frames the establishment of the product advancement process. It records the necessities of a framework as well as has a depiction of its significant highlight. A SRS is essentially an association's seeing (in composing) of a client or potential customer's framework necessities and conditions at a specific point in time (generally) before any genuine configuration or improvement work. It's a two-way protection approach that guarantees that both the customer and the association comprehend alternate's necessities from that viewpoint at a given point in time.
The composition of programming necessity detail lessens advancement exertion, as watchful audit of the report can uncover oversights, mistaken assumptions, and irregularities ahead of schedule in the improvement cycle when these issues are less demanding to right. The SRS talks about the item however not the venture that created it, consequently the SRS serves as a premise for later improvement of the completed item.

3.2 Software and Hardware Require

3.2.1 Hardware System Configuration:

1. Processor - Intel core i3 or greater
2. Speed - 2. 5 Gigahertz
3. RAM - 1 GB minimum
4. Hard Disk - 50 GB minimum

3.2.2 Software System Configuration:

1. Operating System - Windows (7 and greater)
2. Front End - MatLab GUIDE
3. Programming Language - MatLab Script
4. Library - Image Processing Toolbox, Neural Network Toolbox.
5. IDE - MatLab

3.3 System Requirements

3.3.1 Functional Requirements

The Functional Requirements Definition reports and tracks the fundamental data needed to successfully characterize business and practical necessities. The Functional Requirements Definition report is made amid the Planning Phase of the undertaking. Its target group is the undertaking supervisor, task group, venture support, customer/client, and any partner whose data/regard into the necessities definitions procedure is required. They are:

1. Input: ImageNet Dataset.

   Basically ImageNet Is a Project, it's an ongoing research effort to provide a database of image to world which can be easily accessible by the researchers. It has over 10 million images which hand annotated to indicate what objects are present the image.

2. Input: Image.

   Behaviour: The image is preprocessed by removing noise and reshape. Output: A preprocessed image will be extracted.

3. Input: preprocessed image.

   Behaviour: The feature of the input image will be extracted regionprops function calls Output: Feature extracted image.

4. Input: preprocessed image.

   Behaviour: Here the image will be segmented into different parts of object and different object will be extracted. Output: Segmented image.

3.3.2 Non Functional Requirements

1. Reliability: The framework ought to be dependable and solid in giving the functionalities. When a client has rolled out a few improvements, the progressions must be made unmistakable by the framework. The progressions made by the Programmer ought to be unmistakable both to the Project pioneer and in addition the Test designer.

2. Maintainability: The framework observing and upkeep ought to be basic and target in its approach. There should not be an excess of occupations running on diverse machines such that it gets hard to screen whether the employments are running without lapses.

3. Performance: The framework will be utilized by numerous representatives all the while. The framework ought not succumb when numerous clients would be utilizing it all the while. It ought to permit quick availability to every last bit of its clients. For instance, if two test specialists are all the

while attempting to report the vicinity of a bug, then there ought not be any irregularity at the same time.

4. Portability: The framework should to be effectively versatile to another framework. This is obliged when the web server, which is facilitating the framework gets adhered because of a few issues, which requires the framework to be taken to another framework.

5. Scalability: The framework should be sufficiently adaptable to include new functionalities at a later stage. There ought to be a classic channel which can gratify the new functionalities.

6. Flexibility: Flexibility is the capacity of a framework to adjust to changing situations and circumstances, and to adapt to changes to business approaches and rules. An adaptable framework is one that is anything but difficult to configure or adjust because of diverse client and framework prerequisites. The deliberate division of concerns between the trough and motor parts helps adaptability as just a little bit of the framework is influenced when strategies or principles change.

3.4 System Design

Here we are trying to build a image annotation framework using convolutional neural network to recognise a objects in an image and annotate the objects inside the image. This framework will be having to input an image from the user and the user will be having option to select objects in the image either manually or automatically which will then be automatically annotated.

3.4.1 Design Consideration

While doing this framework we considered a simple graphical user interface where user can upload a bunch of images for object recognition, and user will be having a option to detect objects either manually or automatically or live object detection. Usually way of detecting objects in an image takes lots of time where we need to train a model, so we are giving an option to the user to detect object manually

when our image segmentation fails. So this framework will be easy to use than other frameworks in the same field.

3.4.2 System Architecture

Image annotation framework using neural network uses AlexNet. In this framework user will be uploading the image through MatLab GUI, user will be given two option in the GUI whether user wants to detect the objects either manually dragging a rectangle where the user wants to detect the object or will be having a feature to detect the objects automatically where it takes some time to processes the image in the background and it will automatically detect the objects in the image. And also we can detect the objects in the image using live webcam. While designing the system we wanted to develop a framework which is simple and efficiently fast to detect the objects and annotate them.

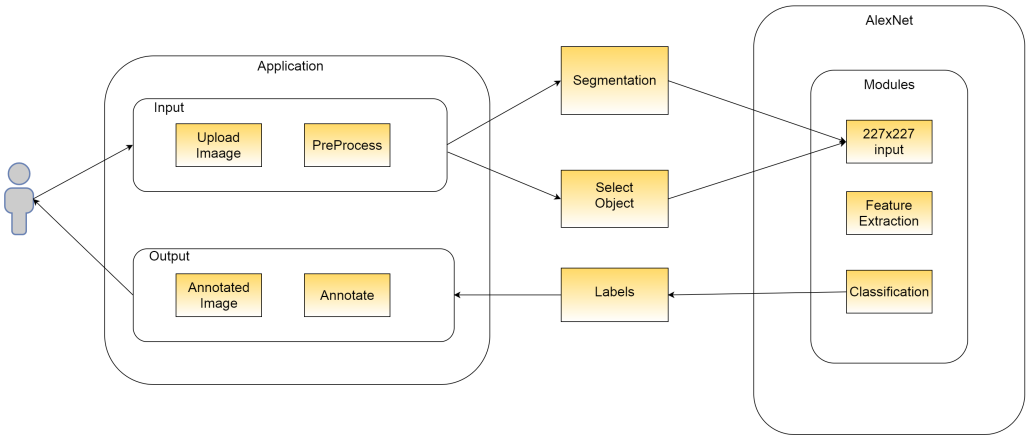

Figure 3.1: System Architecture

3.4.3 Use Case Diagram

A use case chart is of behavioural diagram produced using a case examination. Its article is to show a graphical outline of the handiness gave by a system with respect to entertainers, their targets (addressed as use cases); and some arbitrary conditions between those usage cases. Use case chart gives us the data about how that clients and utilization cases are connected with the framework. Use cases are used amid prerequisites elicitation and examination to speak to the useful of the framework.

Use cases concentrate on the conduct of the framework from an outside perspective. A use case depicts a capacity gave by framework that yields an obvious result for a performer. An use case chart is of behavioral diagram produced using a these cases examination. Its article is to show a graphical outline of the handiness gave by a system with respect to entertainers, their targets (addressed as use cases); and some arbitrary conditions between those usage cases.

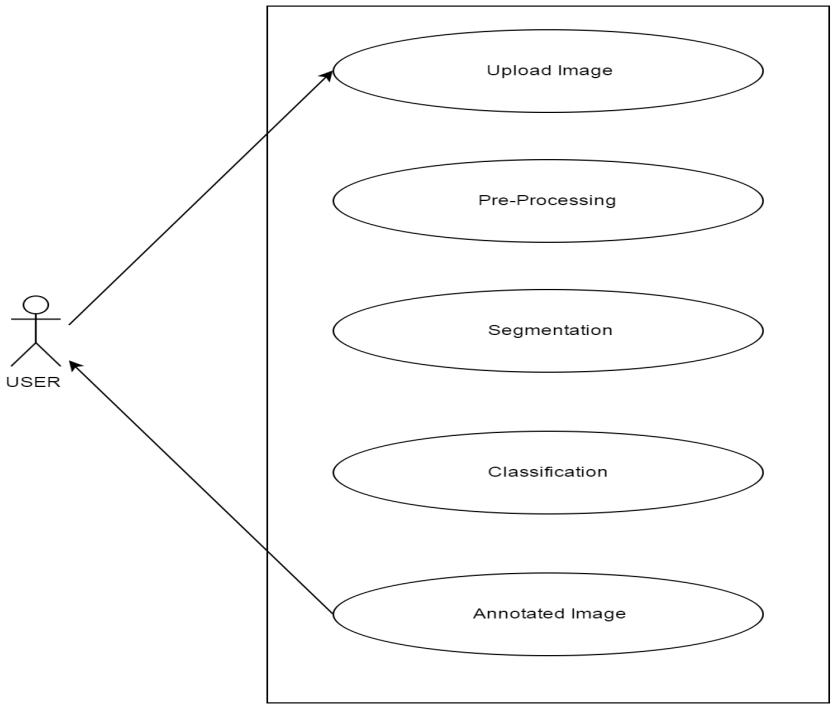

Figure 3.2: Use Case Diagram

3.4.4 Sequence Diagram

A sequence diagram is an integrated Modelling Language is a sort of communication diagram that shows procedures work each other and in what request. Our application begins with the user giving an image as input. Then our application reprocesses the image and communicates with the AlexNet to classify the parts of the segmented image. The classified labels are returned to the application by the AlexNet which are converted into tags and added to form annotated image.

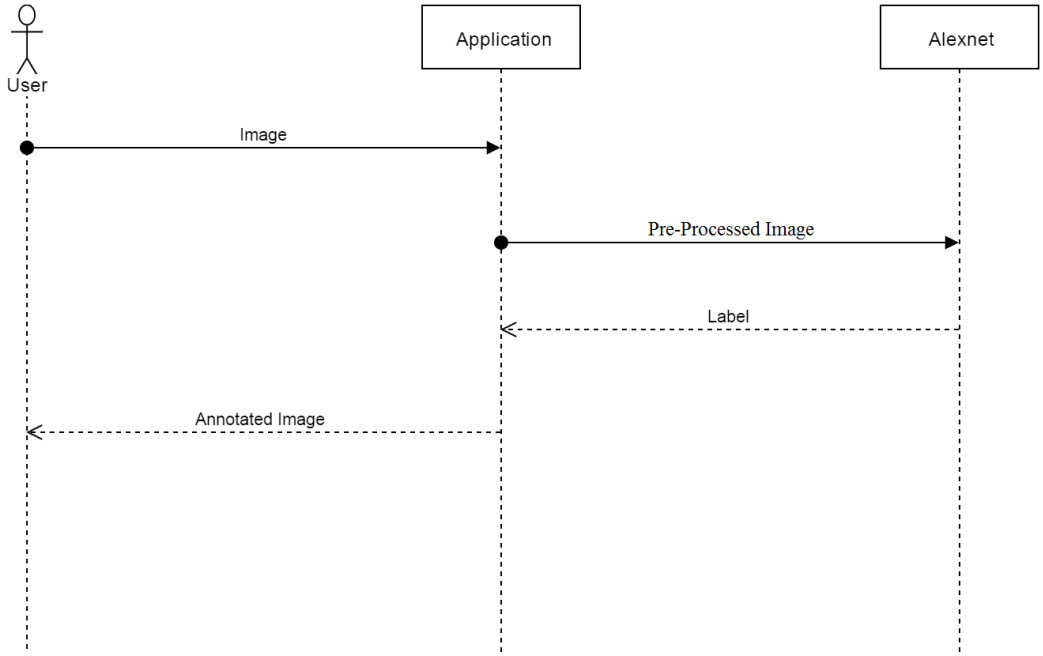

Figure 3.3: Sequence Diagram

3.4.5 DataFlow Diagram

The Data Flow Diagram is clear graphical formalism that can be used, to address a system, to the extent the data to the structure; diverse get ready did on this data and the yield data made by the structure. A Dataflow Diagram model uses an incredibly foreordained number of primitive pictures to address the limits performed by a system and the data stream among the limits.
The Framework gives user a opportunity to upload a image of his choice and the framework will starts pre processing the image, pre processing involves image to conversion to black and white and binary image etc and then image will be segmented and applied to the alexnet neural network model for object recognition and categorization and the annotated image will be presented to the user with the tag.

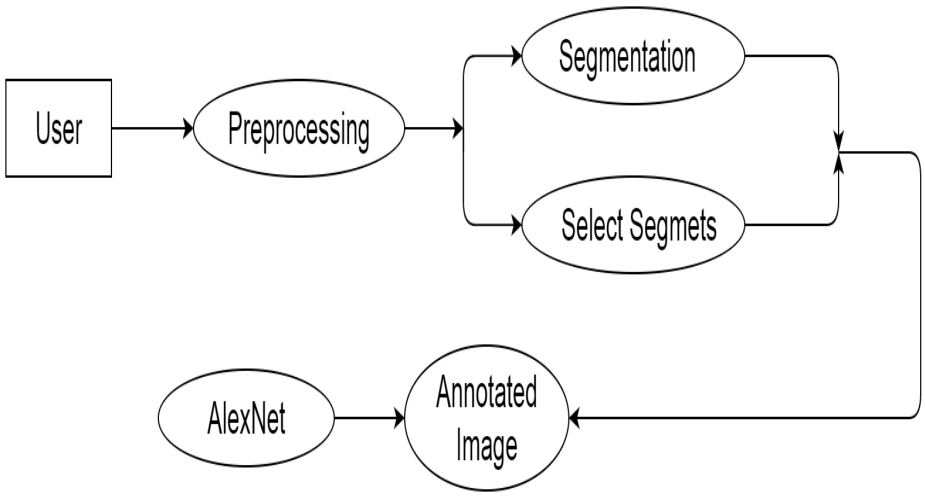

Figure 3.4: Data Flow diagram

Chapter 4

Modelling and Implementation

4.1 Implementation

Now we know the well understood designing of the prototype, it's time to implement the design using the MatLab Software. Implementation of the design calls for the system hardware and software requirements. Before we start implementing the design will be thoroughly understanding the requirement specifications.
The implementation has many steps:

1. Methodical planning.
2. Examination of constraints in system.
3. Assessment of methods and changes in the system.
4. Platform selection.

Execution of any product is continually preceded by basic decisions as for decision of the stage, the lingo used, thus on and so forward these decisions are frequently affected by a few elements, for example, genuine environment in which the framework meets expectations, the velocity that is needed, the security concerns, and other execution choices that was made before the usage of this venture.

They are as follows:

1. Choosing of programming language.
2. Guidelines for coding.

4.2 Coding Platform-MatLab

The MATLAB application is built around the MATLAB scripting language. Common usage of the MATLAB application involves using the Command Window as an interactive mathematical shell or executing text files containing MATLAB code. MATLAB has structure data types. Since all variables in MATLAB are arrays, a more adequate name is "structure array", where each element of the array has the same field names. In addition, MATLAB supports dynamic field names (field look-ups by name, field manipulations, etc.). Unfortunately, MATLAB JIT does not support MATLAB structures, therefore just a simple bundling of various variables into a structure will come at a cost. MATLAB can call functions and subroutines written in the programming languages C or Fortran. A wrapper function is created allowing MATLAB data types to be passed and returned. The dynamically loadable object files created by compiling such functions are termed "MEX-files" (for MATLAB executable). Since 2014 increasing two-way interfacing with Python is being added.

Following are the basic features of MATLAB

1. It is a high-level language for numerical computation, visualization and application development.
2. It also provides an interactive environment for iterative exploration, design and problem solving.
3. It provides vast library of mathematical functions for linear algebra, statistics, Fourier analysis, filtering, optimization, numerical integration and solving ordinary differential equations.

4. It provides built-in graphics for visualizing data and tools for creating custom plots.

5. MATLAB's programming interface gives development tools for improving code quality maintainability and maximizing performance.

6. It provides tools for building applications with custom graphical interfaces.

7. It provides functions for integrating MATLAB based algorithms with external applications and languages such as C, Java, .NET and Microsoft Excel.

4.2.1 MatLab Technical Setup Details

- Software Utilized - MatLab for windows 10
- Software File - MATLAB2017A
- Setup Size- 1478 MB
- Setup sort - Online/Offline .exefile
- Architecture for the similarity - 32 Bit x86/64 Bit x64)
- License-Proprietary commercial software.
- Platform-IA-32, x86-64

4.2.2 MatLab User Interface

MatLab is multi-paradigm numerical computing environment. Implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, C#, Java, Fortran and Python. It will also associate with execution of the .m file.
GUIs (also known as graphical user interfaces or UIs) provide point-and-click control of software applications, eliminating the need to learn a language or type commands in order to run the application.
MATLAB apps are self-contained MATLAB programs with GUI front ends that automate a task or calculation. The GUI typically contains controls such as menus, toolbars, buttons, and sliders. Many MAT LAB products, such as Curve Fitting Toolbox, Signal Processing Toolbox, and Control System Toolbox include apps with custom user interfaces. You can also create your own custom apps, including their corresponding UIs, for others to use.
GUIDE (GUI development environment) provides tools to design user interfaces for custom apps. Using the GUIDE Layout Editor, you can graphically design your UI. GUIDE then automatically generates the MATLAB code for constructing the UI, which you can modify to program the behavior of your app.

4.3 Modules

- Pre-Processing
- Feature Extraction
- Segmentation
- Object Detection
- Image Classification
- Annotation

4.3.1 Preprocessing

Goals:

- It enhances the image features and appearance
- Improves the image facts and suppress the unwanted features like noise.

Image pre-processing is the name for operations on images at the lowest level of abstraction whose aim is an improvement of the image data that suppress undesired distortions or enhances some image features important for further processing. It does not increase image information content. Its methods use the considerable redundancy in images. Neighboring pixels corresponding to one object in real images have the same or similar brightness value and if a distorted pixel can be picked out from the image, it can be restored as an average value of neighboring pixels. Image pre-processing tool, created in MatLab, realizes many brightness transformations and local pre-processing methods.
Preprocessing step can increase the Reliability and it has several methods in it. The different sorts of pre-processing steps are:

1. Image resize
2. Image grayscale conversion
3. Binary conversion
4. Edge detection
5. Noise removal operation
6. Filters
7. Pixel brightness correction

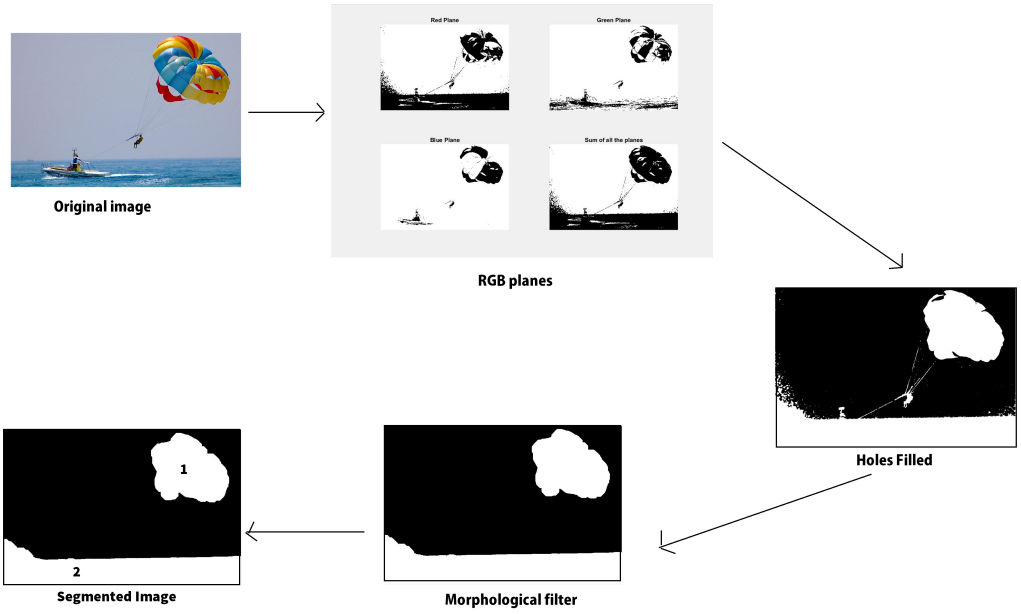

Figure 4.1: Pre processing and segmentation of the image

4.3.2 Feature Extraction

The extraction of image content description is very important. This step as name indicates, it will extract the features which we want from image, which will help in recognize the objects. It will reduce the amount of resources that is needed for describing a large dataset. There are many algorithms for extracting features like SIFT, SURF, HOG, ORB etc.
Morphology is a broad set of image processing operations that process images based on shapes. Morphological operations apply a structuring element to an input image, creating an output image of the same size. In a morphological operation, the value of each pixel in the output image is based on a comparison of the corresponding pixel in the input image with its neighbors. By choosing the size and shape of the neighborhood, you can construct a morphological operation that is sensitive to specific shapes in the input image. The most basic morphological operations are dilation and erosion. Dilation adds pixels to the boundaries of objects in an image, while erosion removes pixels on object boundaries.
RegionProps will find and describe features, which will be stored in form of vectors. RegionProp function is a key point detector and store the result in binary format. So basically, key-points from the object are extracted first and stored in database or in text file. Regionprop function is used to calculate the centriod.
Regionprops measures a variety of image quantities and features in a black and white image. One of these particular properties is the centroid. This is also the centre of mass. You can think of this as the "middle" of the object. This would be the (x,y) locations of where the middle of each object is located. As such, the Centroid for regionprops works such that for each object that is seen in your image, this would calculate the centre of mass for the object and the output of regionprops would return a structure where each element of this structure would tell you what the centroid is for each of the objects in your black and white image. Centroid is just one of the properties.

4.3.3 Segmentation

In computer vision, image segmentation is the process of partitioning a digital image into multiple segments. The goal of segmentation is to simplify and change the representation of an image into something that is more meaningful and easier to analyze.
It is a process of dividing an image in to multiple blocks; it is basically used for locating an object and boundaries in an image. And it assigns labels to pixels in images. Each pixel in divided blocks will have similar attributes.
Techniques used in Segmentation:

- Segmenting Foreground and background

Application of Segmentation:

- Object Detection
- Recognition systems
- Video surveillance, etc.

Thresholding is the simplest method of image segmentation. From a grayscale image, thresholding can be used to create binary images. The simplest thresholding methods replace each pixel in an image with a black pixel if the image intensity is less than some fixed constant T or a white pixel if the image intensity is greater than that constant.
The color thresholding technique is being carried out based on the adaptation and slight modification of the grey level thresholding algorithm. Multilevel thresholding has been conducted to the RGB color information of the object extract it from the background and other objects. Different natural images have been used in the study of color information. The results showed that by using the selected threshold values, the image segmentation technique has been able to separate the object from the background.

4.3.4 Object Detection

AlexNet is one of the deep ConvNets designed to deal with complex scene classification task on Imagenet data. The task is to classify the given input into one of the 1000 classes. AlexNet has 5 convolutional layers, 3 sub sampling layers, 3 fully connected layers. The non linearity used in the feature extractor module of the AlexNet is ReLU. AlexNet uses dropout. A fully trained AlexNet on ImageNet data set can not only be used to classify Imagenet data set but it can also be used without the output layer to extract features from samples of any other data set. ImageNet is a dataset of over 15 million labeled high-resolution images belonging to roughly 22,000 categories. The images were collected from the web and labeled by human labelers using Amazons Mechanical Turk crowd-sourcing tool. Starting in 2010, as part of the Pascal Visual Object Challenge, an annual competition called the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) has been held. ILSVRC uses a subset of ImageNet with roughly 1000 images in each of 1000 categories. In all, there are roughly 1.2 million training images, 50,000 validation images, and 150,000 testing images.

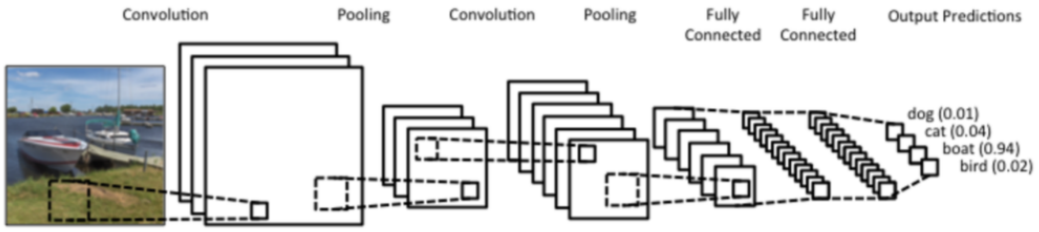

Table 4.1: Object detection and segmentation

4.3.5 Annotation

The output of the last layer of the AlexNet will give the mapped label in the data set which will be used by our framework to add tags to the given image and form an annotated image.

Chapter 5

Testing, Results and Discussion

Testing is an essential stage in the advancement life cycle of the item. This is the stage, where the remaining lapses, if any, from all the stages are identified. Thus testing performs an extremely discriminating part for quality certification and guaranteeing the dependability of the product.
Amid the testing, the system to be tried was executed with a situated of experiments and the yield of the project for the experiment was assessed to figure out if the project was executing of course. Slips were discovered and adjusted by utilizing the beneath expressed testing steps and remedy was recorded for future references.
Consequently, a progression of testing was performed on the framework, before it was prepared for usage. It is the procedure used to help recognize the accuracy, fulfilment, security, and nature of created PC programming. Testing is a procedure of specialized examination, performed for the benefit of partners, i.e. proposed to uncover the quality-related data about the item as for connection in which it is planned to work. This incorporates, however is not restricted to, the procedure of executing a project or application with the goal of discovering lapses.

5.1 Testing Types

A methodology for framework testing coordinates framework experiments and outline systems into very much arranged arrangement of steps that outcomes in

the fruitful development of programming. The testing method must co-work test arranging, experiment configuration, test execution, and the resultant information accumulation and assessment. A procedure for programming testing must suit low-level tests that are important to confirm that a little source code fragment has been effectively actualized and additionally abnormal state tests that are accept significant framework capacities against client prerequisites.

5.1.1 Unit Testing

Unit testing centres check exertion on the unit of programming configuration (module). Utilizing the unit test arrangements, arranged in the configuration period of the framework improvement as an aide, essential control ways are tried to uncover lapses inside of the limit of the modules.
The interfaces of each of the module were tried to guarantee legitimate stream of the data into and out of the modules under thought. Limit conditions were checked. Every single autonomous way was practiced to guarantee that all announcements in the module are executed in any event once and all lapse taking care of ways were tried. Every unit was completely tried to check in the event that it may fall in any conceivable circumstance. This testing was done amid the programming itself. Toward the end of this testing stage, every unit was discovered to be working palatably, as respect to the normal yield from the module.

5.1.2 System Testing

After the Integration testing, the product was totally collected as a bundle; interfacing slips have been uncovered and rectified the last arrangement of programming tests, approval tests start. Here the framework was tried against framework necessity determination. Framework testing was really a progression of diverse tests whose basic role was to completely practice the PC based framework. Albeit every test has an alternate reason all work to confirm that all framework components have been legitimately incorporated and perform distributed capacities.

5.1.3 Integration Testing

Information can be lost over an interface: one module can have an unfriendly impact on another's sub capacities, when joined may not deliver the fancied real capacity.worldwide information structures can exhibit issues. Combination testing was a symmetric method for the building the project structure while in the meantime directing tests to uncover mistakes connected with the interface. All modules are consolidated in this testing step. At that point the whole program was tried in general. Integration testing is another part of testing that is by and large done with a specific end goal to reveal mistakes connected with stream of information crosswise over interfaces. The unit-tried modules are gathered together and tried in little portion, which make it less demanding to confine and right blunders. This methodology is proceeded with unit I have incorporated all modules to frame the framework overall.

5.1.3.1 Table for Integration testing

Test cases	Name of the test	Feature being extracted	Input & Excepted	output	Actual output	status
1	Pre-processing of image.	Convert raw image to grey,binary.	Raw image	Convert to grayscale image and binary image	As expected,grayscale image and binary image	Pass
2	Feature Extraction	From an image, extract features like shape, color, location and texture	Grey image	Feature which extracted and should be stored in any of the format like Mat form	Feature extracted and stored in matrix.	Pass
3	Segmentation	It must segment image and differentiate objects in image by drawing markers	Original raw image, which will be transformed into regions	Marked image	Marked image	Pass
4	Tag	Mapping is done with the existing data set.	Mapping information is retrieves	Pass as input during Training	Passed as input during Training	Pass

5	Annotation	Match the test image with already trained image dataset	Images	Annotated image with tags	Annotated image with tags	Pass

Table 5.1: Integration testing

5.1.4 Performance Testing

The performance testing guarantee that the yield being created inside of as far as possible and time taken for the framework accumulating, offering reaction to the clients and solicitation being sent to the framework keeping in mind the end goal to recover the outcomes.

5.1.5 Validation Testing

The validation testing can be characterized from multiple points of view, yet a basic definition is that.

5.1.6 Black Box testing

In this testing by knowing the inside operation of an item, tests can be led to guarantee that "all apparatuses network", that is the inner operations performs as per detail and all interior segments have been sufficiently worked out.

5.1.6.1 Table Black Box Testing

Test cases	Feature being tested.	Input	Output	Remarks	status
1	Pre-processing & Convert image to grayscale	Original image	Grey image	Without knowing the implementation details expected output was grey image	Pass
2	Pre-processing & Convert image into binary image.	Original image	Binary image.	Without knowing the implementation details expected output was Binary image.	Pass
3	Segmentation	Segment image.	Original image.	It draws a circle around objects with marker	Pass
4	Check for annotation	Pruned tags	Annotated image	List of Tags is displayed	Pass

Table 5.2: Black Box

5.1.7 White Box Testing

This testing is additionally called as glass box testing normally done by code designers. In this testing, by knowing the predefined capacity that an item has been intended to perform test can be directed that shows every capacity is completely operation in the meantime scanning for lapses in every capacity.

5.1.7.1 Table White Box Testing

Test cases	Feature being tested.	Input	Output	Remarks	status
1	Pre-processing & Convert image to grayscale	Original image	Grey image	Expected output was grey image by calling grey( )	Pass
2	Pre-processing & Convert image into binary image.	Original image	Binary image.	Expected output was binary image by calling Binary( )	Pass
3	Segmentation	Segment image.	Original image.	It draws a circle around objects with marker by calling seg function.	Pass
4	Check for annotation	Pruned tags	Annotated image	Tags are displayed.	Pass

Table 5.3: White Box Testing

5.1.8 Acceptance Testing

This is the last phase of testing process before the framework is acknowledged for operational utilization. The framework is tried inside of the information supplied from the framework procurer instead of recreated information. This framework meets the utilitarian prerequisites.

5.2 Result

This area portrays the outcomes got after the execution of actualized re-enactment. The accompanying screen shots characterize the outcomes or yields that are got after orderly execution of all modules of the framework.

This Framework will ask the user to input a image from the folder to annotate the image either by selecting manual mode or Auto mode or user can detect the live object from the web cam feed.

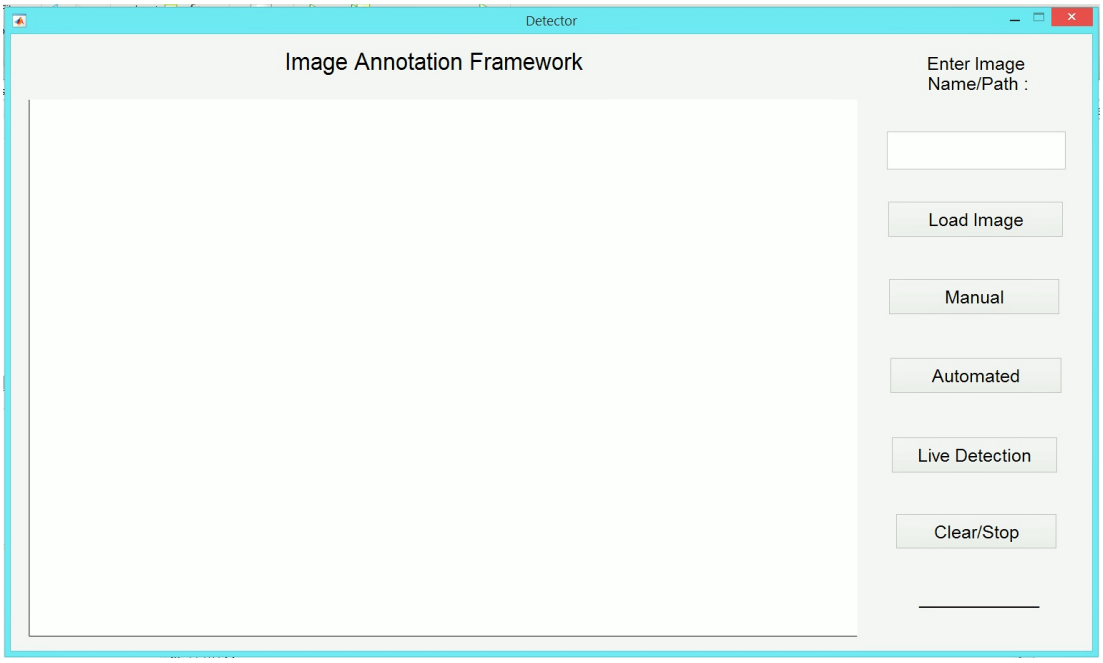

Figure 5.1: Image Annotation Framework

The user interface in the above consists load image,which is used to load the image from the folder. Option for manual annotation,automatic annotation, and also live object detection using webcam.
User Can upload the image by clicking the load image button,and select option from below to annotate the image.

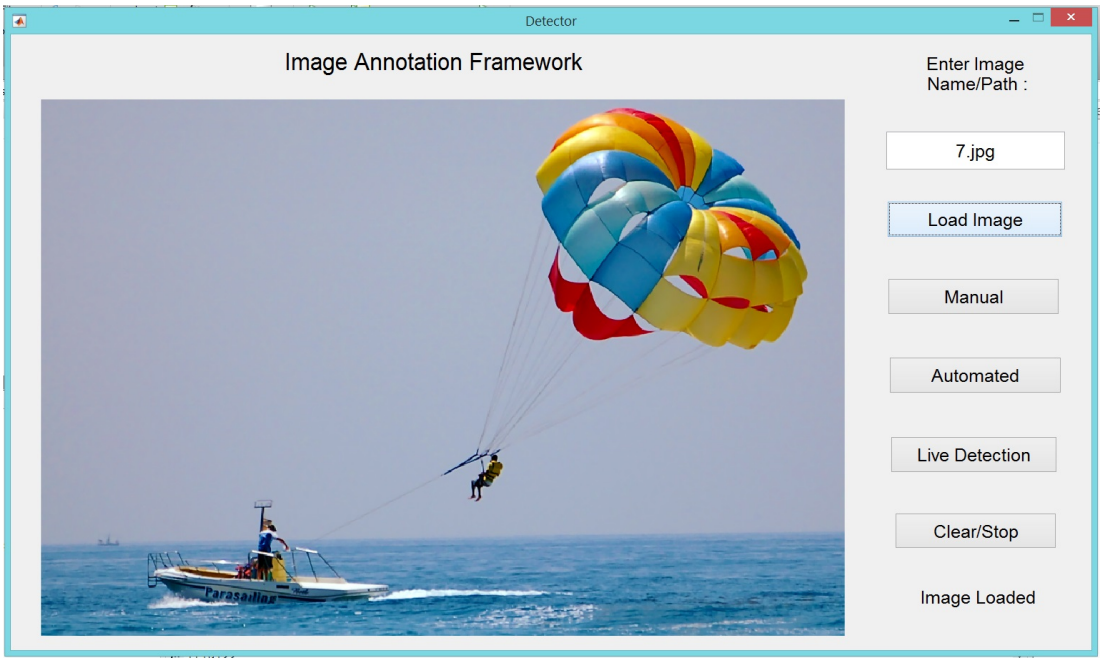

Figure 5.2: Loading the image in the framework

Annotation can be done in two ways either manually dragging box over the image to find the objects or click automatic button to auto annotate the image.

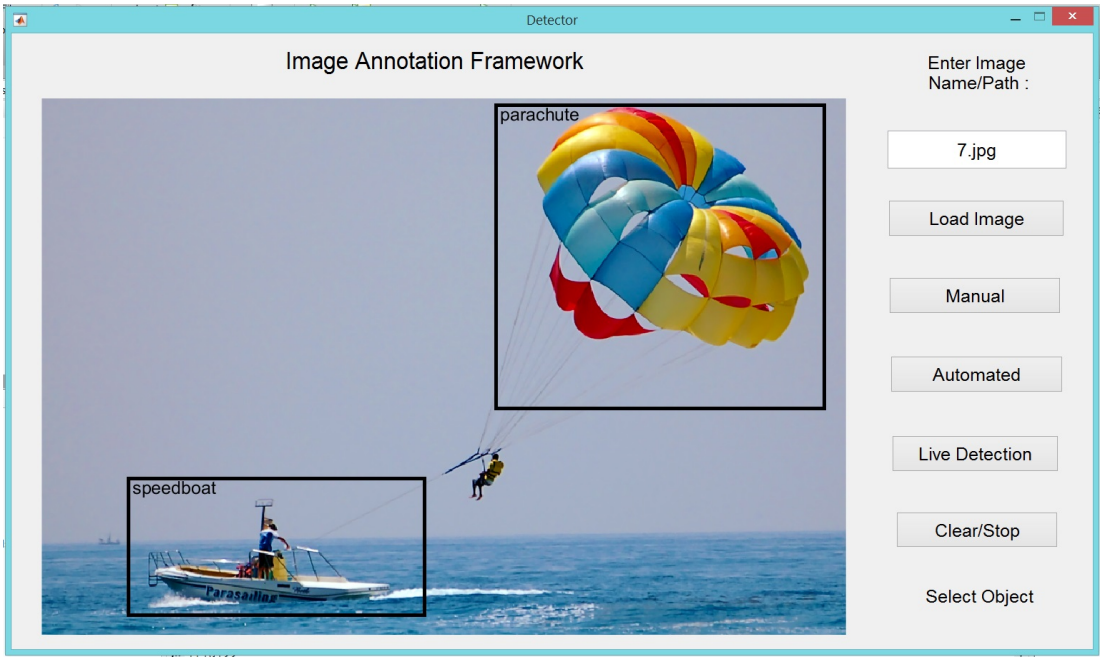

Figure 5.3: Mannual Annotation

Objects in the scene can be detected by using web cam that is live objects can detected.

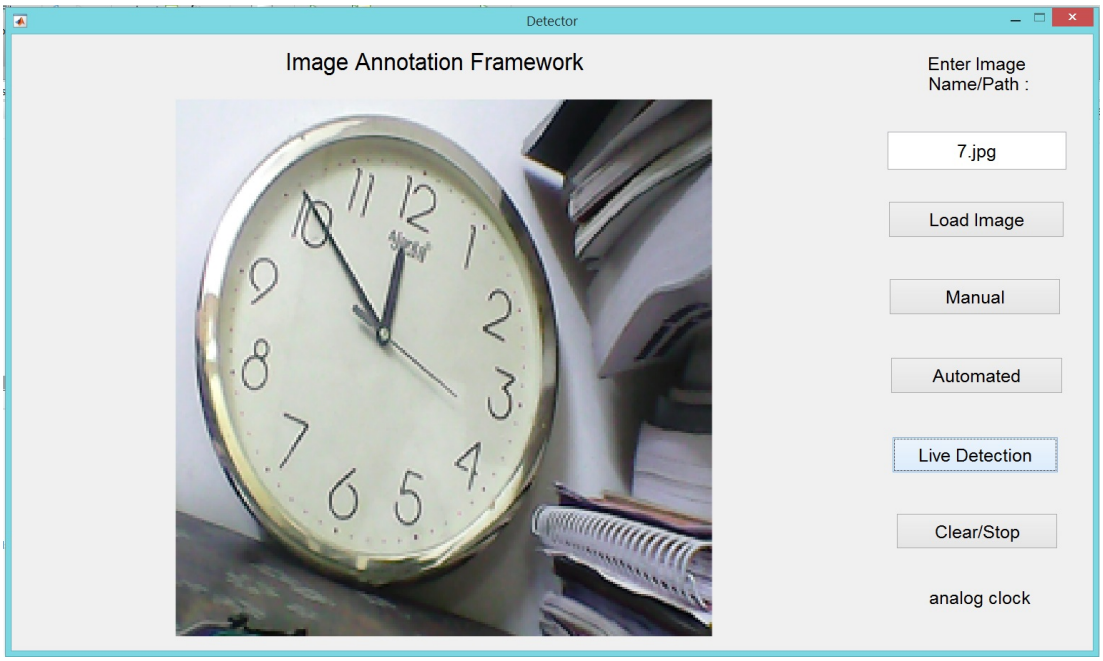

Figure 5.4: Live object Detection

Object is detected by the framework successfully, which says its a Analog clock

Objects in the figure detected is computer keyboard.

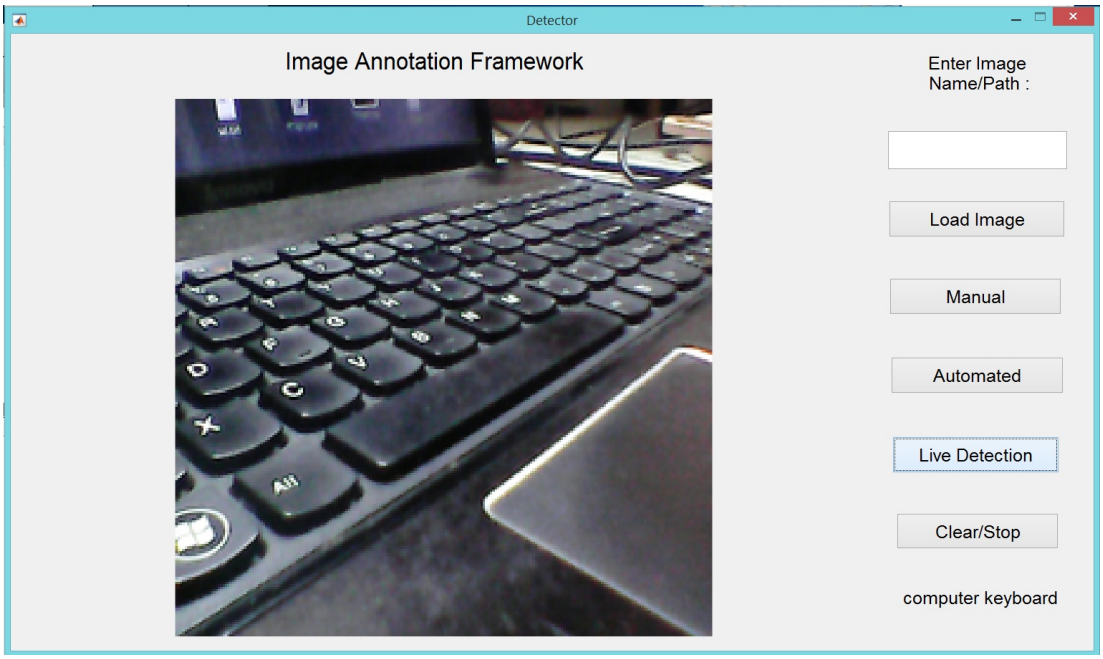

Figure 5.5: Live object Detection

Chapter 6

Conclusion and Future Work

6.1 Conclusion

Here a novel framework is proposed for annotating the images using ImageNet as Dataset. For every image in the dataset, Annotations are produced and are displayed. When we input an image, it will be segmented to detect objects by Image Retrieval Algorithm Based on Convolutional Neural Network. And the input image is compared to correctness of the objects detected.The proposed method gives us a robust methodology for extracting different object of images at low time complexity.We have tested our result with other methods such "Image Classification with Bag of Visual Words" which uses K means clustering technique to classify the image which takes lot of time and space. Our proposed method takes comparatively less time than the other method to detect an object in the image.This mechanism can be used for various purposes in the field of Object detection.

6.2 Future work

Future work to improve the accuracy of the system can take many directions. First, the incorporation of 3-D information in the learning process may improve the models, perhaps through learning via stereo images or 3-D images. Additionally, shape information can be utilized to improve the modeling process. Second, better and larger amounts of training images per semantic concept may produce more robust models. Contextual information may also help in the modeling and annotation process. Third, this method holds promise for various application domains, including bio-medicine. Finally, the system can be integrated with other retrieval methods to improve usability in the field of computer vision.

Bibliography

1. Jia-Yu Pan, Hyung-Jeong Yang, Pinar Duygulu and Christos Faloutsos,Automatic Image Captioning,2104.

2. Jeon J, Lavrenko V, Manmatha R, Automatic image annotation and retrieval using cross-media relevance models, Jul.2003.

3. Ilaria Bartolini and Paolo Ciaccia,Imagination: Exploiting Link Analysis for Accurate Image Annotation.2012

4. Chong Wang, David Blei, Li Fei-Fei,Simultaneous Image Classification and Annotation,2012.

5. Yashaswi Verma and C. V. Jawahar,Image Annotation Using Metric Learning in Semantic Neighbourhoods,2012

6. Yasuhide MORI, Hironobu and Ryuichi OKA, Image-to-word transformation based on dividing and vector quantizing images with words,1999.

7. Duygulu, P., Barnard, K., Freitas, J., Forsyth, D.A., Object recognition as machine translation: learning a lexicon for a fixed image vocabulary".

8. Guillaumin, M,Ferrari, V. Large scale knowledge transfer for object localization in ImageNet,2012

9. Jeon J, Lavrenko V, Manmatha R, Automatic image annotation and retrieval using cross-media relevance models, Jul.2003.

10. Shaoting Zhang, Junzhou Huang, Hongsheng Li, and Dimitris N. Metaxas , Automatic Image Annotation and Retrieval Using Group Sparsity, IEEE June 2012

11. Rong Jin, Joyce Y. Chai, Luo Si, Effective Automatic Image Annotation Via A Coherent Language Model and Active Learning, ACM, 2004.

12. Karen Simonyan, Andrew Zisserman,Very Deep Convolutional Networks for Large-Scale Image Recognition,2014.

13. Krizhevsky, A., Sutskever, I. and Hinton, G. E,ImageNet Classification with Deep Convolutional Neural Networks.NIPS 2012.

14. Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, Trevor Darrell,DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition,2013.

15. Richard Socher, Brody Huval, Bharath Bhat, Christopher D. Manning, Andrew Y. Ng, Convolutional-Recursive Deep Learning for 3D Object Classification.April 2014.

16. Ronan Collobert, Ronan Collobert , A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning.

17. Collobert, Ronan; Weston, Jason, A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning.

18. Pedro Ballester and Ricardo Matsumura Araujo, On the Performance of GoogLeNet and AlexNet Applied to Sketches.

19. Hailong Liu, Baoan Li, Xue Qiang Lv and Yue Huang,Image Retrieval Algorithm Based on Convolutional Neural Network, august 2016.

20. Ando, R. and Zhang, T, A framework for learning predictive structures from multiple tasks and unlabeled data. 2005.