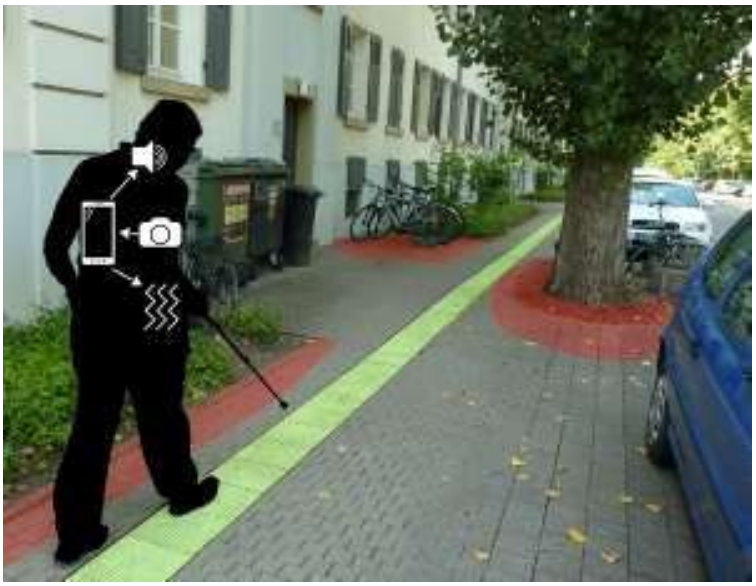
ENVISAGER: SIGHT FOR BLIND AND VISUALLY IMPAIRED

Project Phase - 3



Submitted on **24 April 2017**

By **Team 7 (Team Supreme)**

Nandanamudi, Jyothi kiran (27)
Gatiganti,Sai Venkatesh (9)
Gumma,Madhuri (13)
Nallamotu,Naveena (26)

I. Introduction:

In this document, we intend to provide our objective, significance and features of the project. We also discussed our goals and objectives for the application named "Envisager" and what motivated us for choosing this application. We aim to incorporate some added features to make this application notable.

The main objective of the project is to help the blind and visually impaired people crossing roads by detecting the object nearby and other obstacles by using the smartphone. The smartphone is used to capture the video of surroundings.

II. Project goal and objectives:

1. Significance:

In this hustle and bustle of modern life, even people with no disabilities find many things as a hindrance for completing their activities and daily routines. However, people who are visually disabled are facing many difficulties than the normal people without any disability. As we know, the social constructs are not always designed by keeping them people in mind. They do need support in crossing roads, detecting the object nearby and other obstacles. Even though they use canes, they cannot identify objects above their waists, which is why there is a need a smart way of identifying the object they come across.

With the same motivation, companies like Siemens started building application that would help visually impaired people to navigate through busy roads by considering the GPS and identifying the persons location and help them to reach their location. For the past, few years there are many inventions like smart canes that detect object impact and electronic glass (eSight) that lets the people see but, these items costs a lot and not all can afford to have one.

However, with an android app like ours, people with smartphone can have it and we could help most of them to perform their daily activities without running into any trouble. Our main goal is to develop an application which would allow the blind people to take the pictures through camera and then our system would detect the image captured and identify the object and give them the audio reply describing the object or the naming the object etc.,. Moreover, all the features can be used without spending a dime and all they required is to install the application.

Objectives:

- **a.** To develop a user-friendly application.
- **b.** To provide a smart application that would detect the object accurately.
- **c.** To notify the user about the object with the voice that the person can understand.
- **d.** To test the time taken by the process in different methods (using Clarifai API, Spark Machine Learning, Deep Learning) and choose the best method for our application.
- **e.** To develop an application that can be operated with a minimal cost.
- **f.** To develop a scalable project

2. **Features: Use Case/Scenario**

This project has two different feature (a) Image to text conversion (b) Text to speech conversion.

(a)	Android App	Built an android app which takes continuous frames and gives back annotations .
(b)	Image Classification	Used tensorflow with SoftMax classification algorithm to classify the test image into predefined classes (labeled classes)
(b)	Text to Speech Conversion	Android inbuilt text to speech conversion API is being used

Table 1: Features of the project phase-3

III. Approaches

In this Phase-3 we generated a model to classify the test image to a defined class (Supervised learning model) using tensorflow and an android application to take images surroundings.

Input: Image
Output: Labelled class of an image and Audio (Image Explanation)

1. **Data Source**

ImageNet Dataset: ImageNet dataset has been one of the popular datasets with thousands of images in each node with hierarchy. Even though there are only nouns available for now, there is almost thousand images per node. The database is organized according the hierarchy used by the WordNet.

Our model is trained on the ImageNet Large Visual Recognition Challenge dataset which is subset of ImageNet .some of the classes are "Zebra", "Dalmatian", and "Dishwasher".

2. **Analytic Tool**

Pycharm, Android Studio

3. **Expected Inputs and Outputs:**

Input	Output
Jpeg Image	Class of the image belongs to, audio saying the class name

Table 2: Expected input and the outputs

4. **Algorithms:**

1. SoftMax classification algorithm: To classify images for classes

IV. Related Work:

Clarifai API: Clarifai API accepts a video, image or an audio as an input from the Android application and it will break down the media to analyse it and to use the information to configure and extract the words the media may contain. The API uses machine learning techniques which would improve with time and it gives the output to perform any further actions

Spark: Apache Spark is an open source platform for scalable MapReduce computing on clusters. The main goal of the Spark framework is speeding up Image classification by parallelizing the machine learning to a high-performance computing cluster. We integrate open source SDK library to Spark API. To classify the image/video which is taken as the input from Android APP we use random forest classification algorithm in Spark.

V. Application Specification

1. System Specifications

   a. Software Architecture Diagram

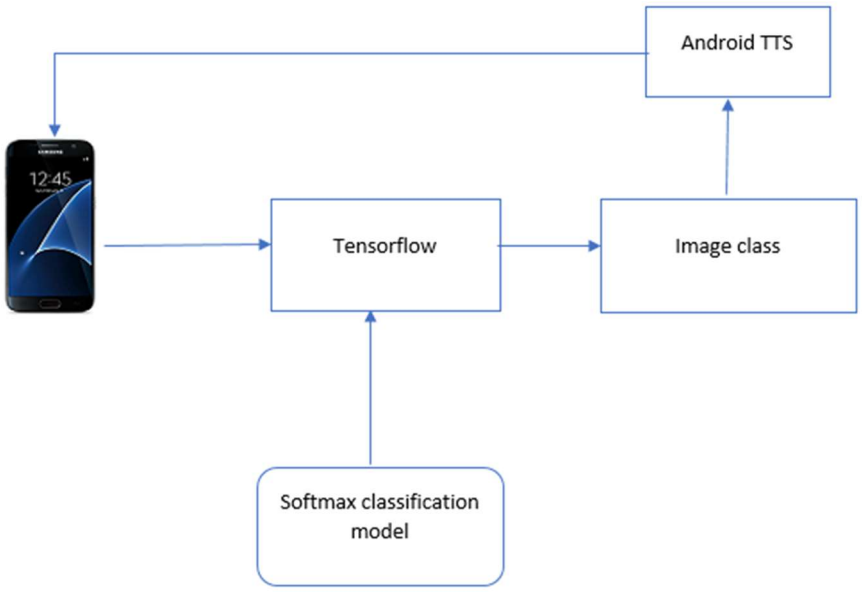


Figure 1: Software Architecture Diagram using Spark AP

b. Inception Module:

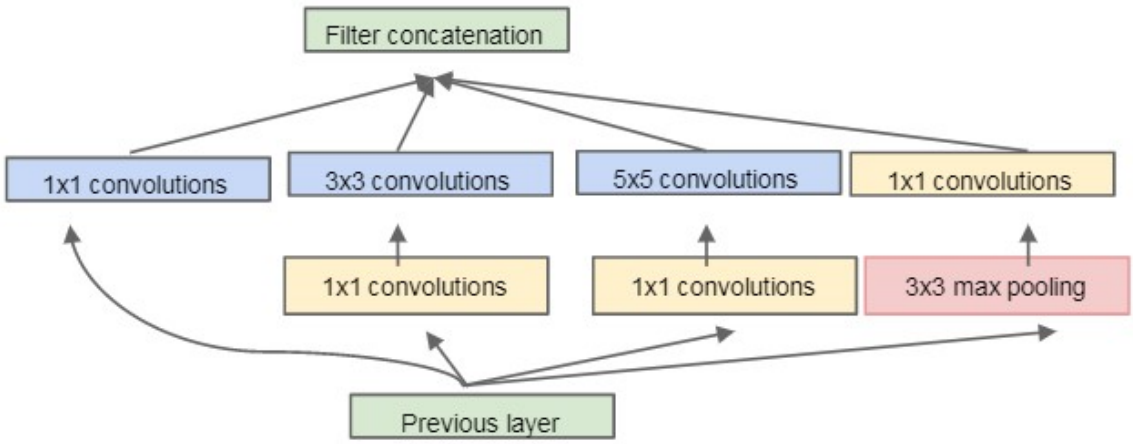


Figure 2:Inception Module

c. Activity Diagram:

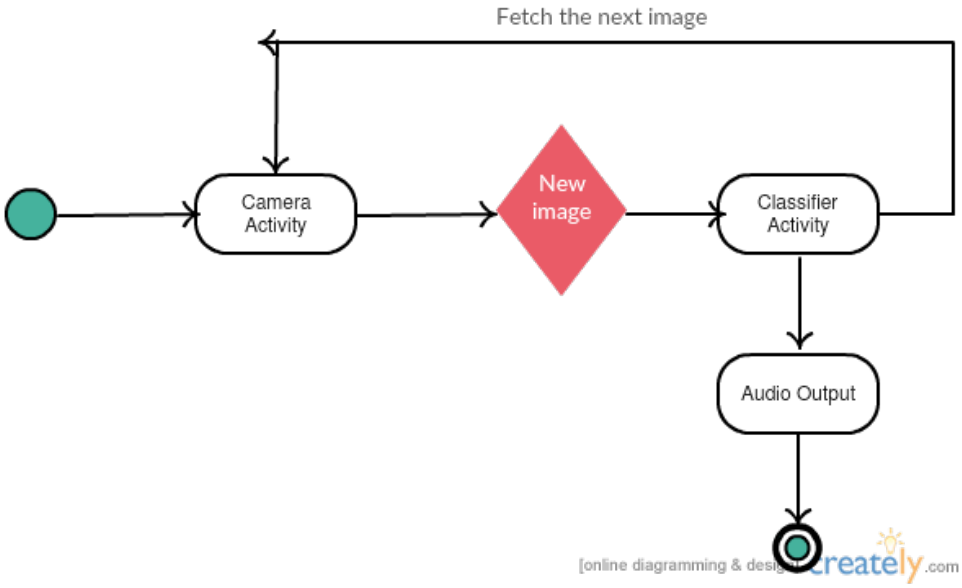


Figure 3: Activity Diagram using tensorflow

d. Sequence Diagram

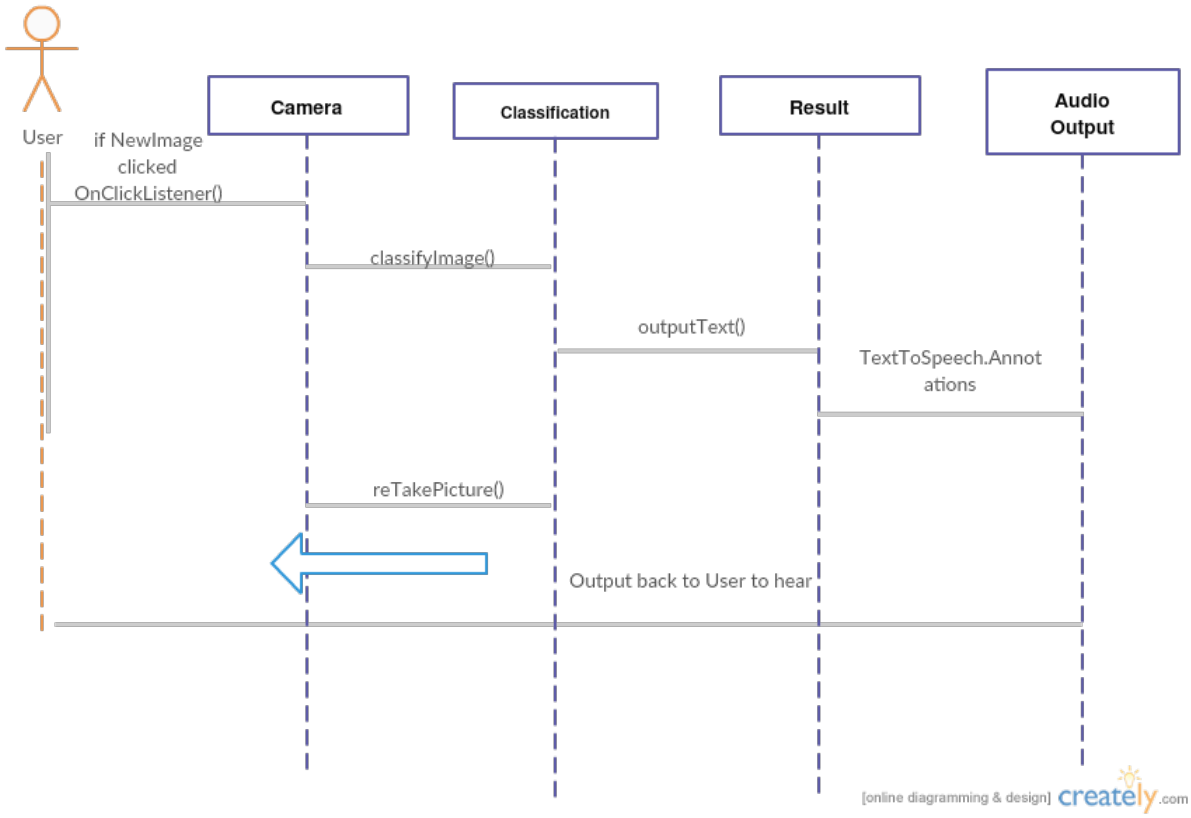


Figure 4 : Sequence Diagram using tensorflow

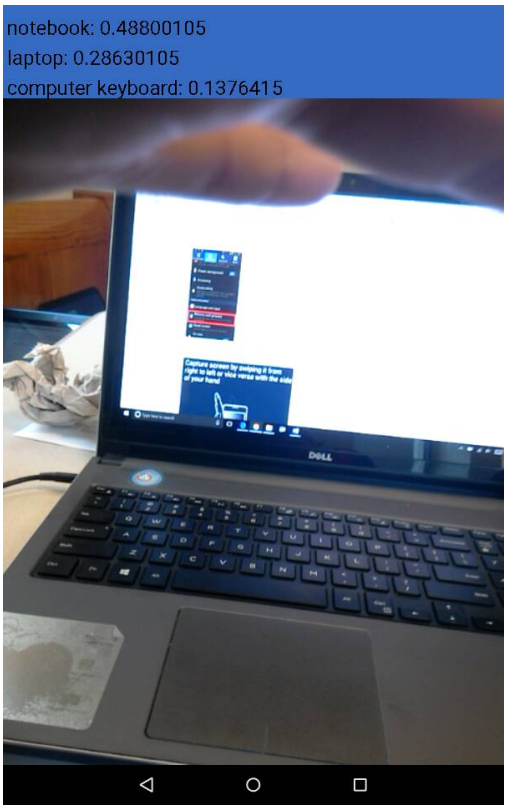
VI. Implementation

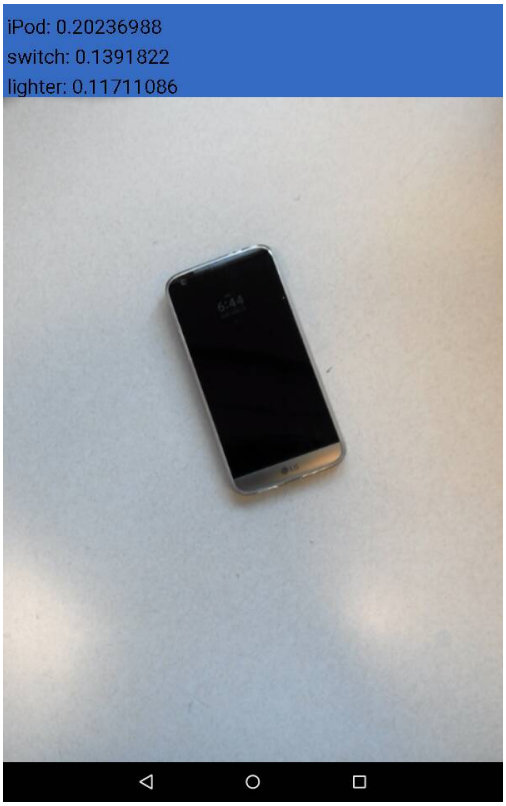
Convolution Neural Networks: The typical architecture of convolutional neural networks is the union of convolutional layers(with normalization and max pooling layers followed by one or more fully connected networks. Often these architectures proved to give best results for various datasets such as MNIST, CIFAR and even ImageNet classification challenge .

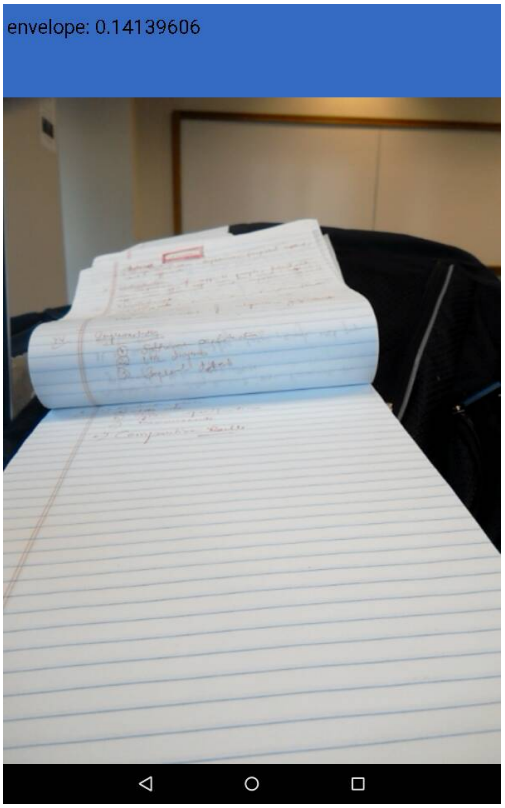
Inspired by the primary visual cortex of the neuroscience model, to handle multiple scales filters of different sizes were used similar to inception model and in the Envisager (proposed model). Similar to inception model all the filters in all the layers are learned and theses inception layers were repeated many times resulting in a 22-layer model.

In the object detection, we used a model proposed by Girshick et al. but instead used multi box prediction rather than single bounding box.

VII. Documentation:







Figures 1-4:Image classification done on sample images

VIII. Project Management:

Implementation status report

Work Completed:

Percentage Contributed	Team Member	Work Done
25%	Venkatesh Gatiganti	Research paper, model training, camera activity and layout
25%	Jyothi Kiran Nandanamudi	Research paper, image classification, classification activity and class, result view
25%	Madhuri, Gumma	Research paper, report
25%	Naveena Nallamothu	Research paper, Data collection, result activity , layouts

Table 3: Work Completed and Contribution

Project URL:

https://github.com/nandanamudi/Big-Data-Analytics-and-Application---Envisager

IX. Reference:

1. https://developer.android.com/training/basics/firstapp/index.html
2. https://www.tutorialspoint.com/android/
3. https://classroom.udacity.com/
4. Aditya Bhaskara, Sanjeev Arora, Rong Ge, and Tengyu Ma. Provable bounds for learning some deep representations. CoRR, .
5. Y. LeCun, J. S. Denker, B. Boser, D. Henderson,, W. Hubbard, R. E. Howard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. Neural Comput .
6. Shuicheng Yan and Min Lin, Qiang Chen. Network in network. CoRR, .