Real Time Object and Color detection

Aqeela Sultan

Ayesha Nawaz

DEPARTMENT OF COMPUTER SCIENCES
COMSATS UNIVERSITY ISLAMABAD,
ATTOCK CAMPUS – PAKISTAN

SESSION 2017-2021

Real Time Object and Color detection

Undertaken by

Ayesha Nawaz

CIIT/FA17-BCS-088/ATK

Aqeela Sultan

CIIT/FA17-BCS-072/ATK

Supervised By:

Miss Sadia Ejaz

A DISSERTATION SUBMITTED AS A PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF BACHELOR OF COMPUTER SCIENCE

DEPARTMENT OF COMPUTER SCIENCES
COMSATS UNIVERSITY ISLAMABAD,
ATTOCK CAMPUS – PAKISTAN

SESSION 2017-2021

UNDERTAKEN

We certify that this is my/our own work. The work has not, in whole or in part, been presented elsewhere for assessment. Where material has been used from other sources it has been properly acknowledged. If this statement is untrue, we acknowledge that we will have committed an assessment offence and shall be liable to punishable action under the plagiarism rules of HEC.

Aqeela Sultan FA17-BCS-072	Ayesha Nawaz FA17-BCS-088

Dated: __________ Dated: __________

FINAL APPROVAL

Certified that we have read this project report submitted by Miss (Aqeela Sultan, Ayesha Nawaz) and it is, in our judgment, of sufficient standard to warrant its acceptance by Department of Computer Science, University of Islamabad, Attock Cantt, for the (BS/MSc degree) in Computer Science.

Committee:

1. External Examiner ____________________________
(Examiner Name)
Designation
University Name

2. Supervisor ____________________________
(Supervisor Name)

3. Chairperson ____________________________
(Chairperson Name)

4. Dean/Director ____________________________
(Dean/Director Name)

DEDICATION

To my Loving Parents and Teachers

Acknowledgement

All praise is to Almighty Allah who bestowed upon us a minute portion of his bound less knowledge by virtue of which we were able to accomplish this challenging task. We are greatly indebted to our project supervisor "Miss Sadia Ejaz". Without her personal supervision, advice and valuable guidance, completion of this project would have been doubtful. We are deeply indebted to them for their encouragement and continual help during this work and we are also thankful to our parents and family who have been a constant source of encouragement for us and brought us the values of honesty & hard work.

Aqeela Sultan Ayesha Nawaz

-------------------------- --------------------------

PROJECT BRIEF

PROJECT NAME	REAL TIME OBJECT AND COLOR DETECTION
ORGANIZATION NAME	COMSATS UIVERSITY ISLAMABAD, ATTOCK CAMPUS
OBJECTIVE	TO HELP USER TO OVERCOME VISUAL DISABILITIES, SAVE EFFORTS, RESOURCES, MONEY AND ALSO FOR LEARNING PERCEPTIVE
UNDERTAKEN BY	AQEELA SULTAN, AYESHA NAWAZ
SUPERVISED BY	MISS SADIA EJAZ COMPUTER SCIENCE COMSATS UIVERSITY ISLAMABAD, ATTOCK CAMPUS
STARTED ON	SEPTEMBER 2017
COMPLETED ON	MAY 2021
COMPUTER USED	HP (ELITEBOOK 8470P)
SOURCE LANGUAGE	PYTHON, JAVA, XML

OPERATING SYSTEM	WINDOW 10
TOOLS USED	ANDROID STUDIO 4.0.1, IDLE (PYTHON 3.6)

ABSTRACT

A model based on scalable object detection using deep neural networks approaches employ convolutional neural networks (CNNs) techniques to identify objects in real-time camera preview. Object detection has the same effect as other computer vision techniques, such as image recognition and image separation, allowing us to understand and analyze scenes in images or videos. The key intention is to dispose the need of dedicated gadgets, different wearable devices and also multiple use of android application. We are imposing this undertaking for getting to know attitude and occasional imaginative and prescient people, the use of android Application with the help of smart phones. It is implemented in android application and it is easy to use in mobile phone or any other smart device. By the use of machine learning algorithm and model, image processing, object detection, speech synthesis the highlighted key additives to use this application. Detecting item, the usage of machine learning and image processing technique can be used in more than one business as well as asocial programs. Purpose of this project is object and color detection for helping color blind people and learning perspective by using speech synthesis and feature extraction in camera view.

Table of Content

Table of figures

Chapter 1

INTRODUCTION

1 Introduction

Object detection is utilized to portray an assortment of related PC vision assignments, which include exercises such as recognizing objects in image. Object recognition is a largely forthcoming scope for detecting object using machine learning methods which are used for social and industrial applications. Color is a component that has been intensely exploited in computerized picture handling, as it is a useful asset that regularly encourages the order and distinguishing proof of items, which can be segregated dependent on the huge number of calculable shading tones. One problem in image processing is related to the recognition and division of features of high color intensity, predominantly identified with essential tones and their supplement, which are utilized in characterizing shading spaces, since there are explicit applications, in which the objects of interest can be effectively recognized because of their high immersion in one RGB shading parts (red - green - blue). There are several techniques used to detect features. The Android application is a product application that sudden spikes in demand for the Android platform. Since the Android stage is expected for PDAs, the ordinary Android application is planned for a wireless or tablet running the Android OS working system. Android provides a comprehensive application framework that allows us to create modern apps for mobile devices in a Java language environment. Today, smartphone users are making the leap. It is estimated that the late 80% of the populace utilizes cell phones. What's more, they utilize their cell phones to meet their day by day needs. Purpose of this project is object and color detection for helping color blind people and learning perspective by using speech synthesis and text labeling in image and in camera view. Thus, we conclude that object detection is useful in day to day life for learning and low visually impaired users.

1.1 Brief

The motive of the object and color detection is to recognize and locate (localize) all known objects in a scene. Helping others and providing them ease in effective way is the thing that motivated us to design such an android app named “Real Time Color and Object Detection”. Computer vision has been around for decades and has many different uses in the real world. Examples include facial recognition, disease/diagnostic detection and real-time object detection (which is our main focus for this article). The whole concept of computer vision depends heavily on machine learning, especially in today's world, and how new technologies are emerging. When a machine aims to categorize objects within an image or webcam in real time, it must train tagged data. This data tagged in object detection is self-tagged maps. Before training the model, we must have many tagged images of specific elements. This means that after training the models, the goal is to identify the items tagged with them in real time. Now, when we actually train the program, we're trying to learn it by learning to monitor. Because it is already tagged images (output), it learns mainly through trial and error, constantly trying to predict the best results. An important component during model training is the loss function, which is a permanent measure of model accuracy. The goal of training is to improve accuracy, so detect damage and damage errors as much as model trains. This is because the machine, like us humans, is constantly learning to make better decisions to reduce the error. A convolutional neural network (ConvNet/CNN) is a deep learning calculation that can take in a data image, give importance (learned loads and predispositions) to various points/objects in the image, and have the option to isolate from each other. The necessary pre-handling in a ConvNet is much lower when contrasted with other layout calculations. While in crude techniques the channels are hand-designed, with sufficient preparation, ConvNets can become familiar with these channels/attributes. ConvNet engineering is like the example of correspondence of neurons in the human brain and is awakened by the Visual Cortex guideline. Solitary neurons respond to updates simply in a limited region of the visual field known as the responsive field. A meeting of these fields cover to cover the entire obvious region.

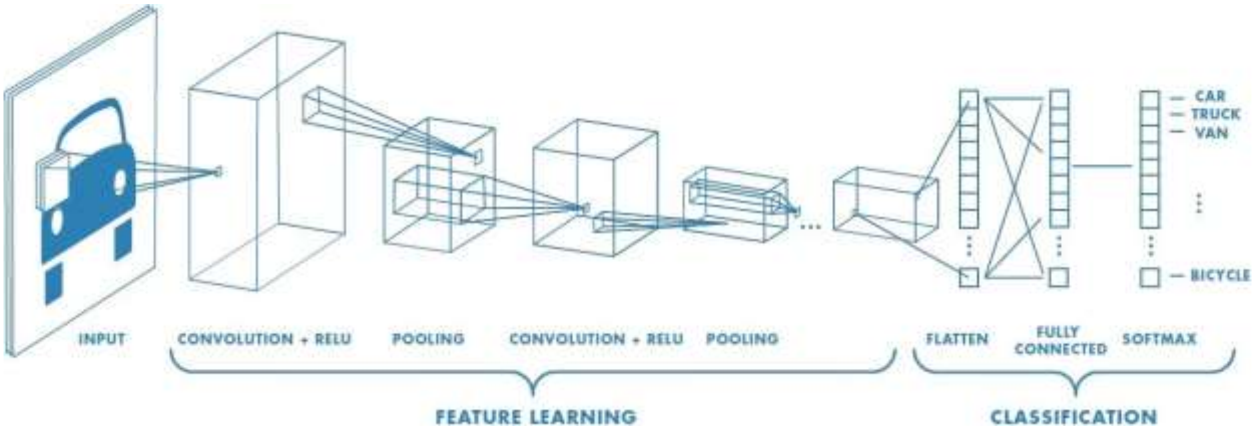


Figure 1-1 CNN Architecture

A neural association involves a couple of remarkable layers, for instance, a data layer, at any rate one covered layer, and a yield layer. They are best used in thing disclosure to see models, for instance, edges (vertical/level), shapes, tones, and surfaces. Costumed layers are convolutional layers in this type of neural association that probably go as a channel that gets entries first, changes over it with a specific model/incorporation, and sends it to the accompanying layer. With more convolutional layers, each time another data is sent off the accompanying convolutional layer, it is changed in a surprising way. On a very basic level, as a truly expanding number of layers of the data experience, the additionally bewildering models that the future can discover can be found. The growing need to run conventional neural network (CNN) models on portable devices with limited computing power and memory resources is encouraging studies of efficient model design. Several effective structures have been proposed in recent years, for example, Mobile Net, Shuffle Net, and Mobile Net V2.

Android utility is software that runs on android platform. Because the android software is built for cell devices an average android utility is designed for smartphones or pill laptop that runs on android OS. Android utility is very easy, handy, portable and fantastic in use. Also because of its characteristic and operation capabilities more capabilities might be achieved. It is a completely effective and upcoming platform.

Speech synthesis which is known as synthetic manufacturing of human voice. Synthesized speech can be created through concatenating portions of recorded speech that is saved in database. This is generated after facts extraction and detection in speech synthesis. The output might be in shape of speech as a way to deliver the impaired users a concept of object and color detected.

Our aim is to develop an application which can be very helpful for color blind persons or for low sight persons. Previously major’s work has been done for this field but we are here to give a project which is a full package. It’ll have all of the features including audio synthesis. Real time image metering will be used in our project which will produce an obstacle detection. The object will have a color so the user will have the choice if he wants to read the color name or wants to hear the name of color. Our project will be able to judge object, then it’ll tell the color of surface.

1.2 Relevance to Course Modules

1.2.1.1 Android development

Android Studio is the IDE that provides Google to develop professional Android applications. It is used to develop a different variety of applications for the Android operating system. It is an IDE & platform to design a user-friendly interface by drag and drop.

1.2.1.2 Machine Learning

In this course we have studied about different algorithms which are used to train models to perfume task automatically after getting the training data.

1.2.1.3 Report Writing Skills

This course is about learning how to write reports and other formal documentation, and, in our project, we need to write our documentation, so this course is helping a lot in this task.

1.2.1.4 Software Engineering

Software engineering is the systematic application of engineering methods in software development.

1.3 Project Background

The idea behind this project particularly occurred in our mind when we think about the low vision and color blind persons. Our aim is to develop an application which can be very helpful for color blind persons or for low sight persons. Previously major's work has been done for this field but we are here to give a project which is a full package. It'll have all of the features including audio synthesis. Real time image metering will be used in our project which will produce an obstacle detection. The object will have a color so the user will have the choice if he wants to read the color name or wants to hear the name of color. Our project will be able to judge object, weather it is a laptop or register, and then it'll tell the color of surface.

1.4 Literature Review

With a specific end goal to influence the task, to yield appropriate output, it is important to experience the past looks into and procedures. This heading contains the complete description of the past work and reports that have been completed in such manner. Different approaches have been used to detect object. We illustrate how these approaches differ from each other, bring to light the merits of each and limitations.

1.4.1 Existing Methods

1.4.1.1 R-CNN

To evade the issue of choosing a huge number of places, Ross Girshick and others proposed a technique in which we used the particular search of only 2000 districts of the image and he called them area recommendations. So, instead of trying to organize the gigantic number of areas, you can simply work with 2000 places. These area 2000 recommendations are produced by using the calculation of the particular game that is composed below.

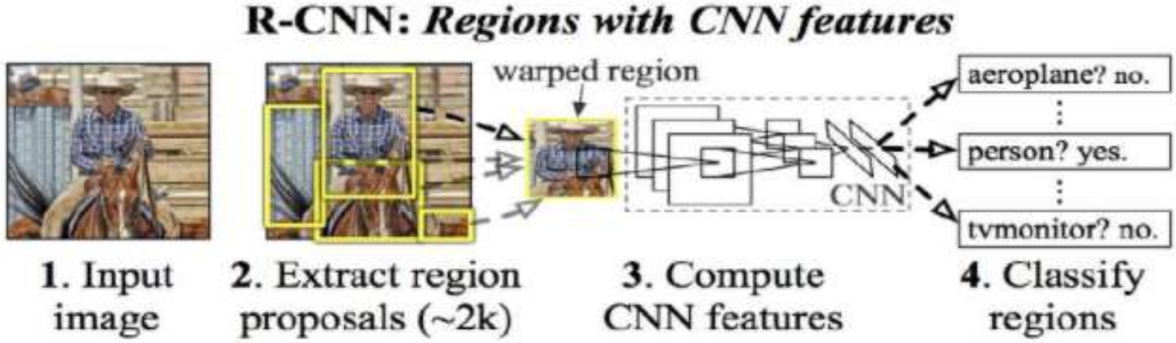


Figure 1-2 R-CNN Architecture

These competing districts of 2000 that are recommendations are twisted into a square and deal with a convolutional neural organization that creates a vector of elements of 4096 dimensions as performance. CNN assumes a role of highlighting extractor and the thick layer of performance comprises the highlights of the image and the separate highlights are addressed in an SVM for the provision of the presence of the article within that competitive district proposal. However, to provide for the presence of an element within the locale proposal, the calculation also predicts four qualities that are counteracted qualities to extend the accuracy of the rebound box. For example, given the locale proposal, the calculation may have anticipated the presence of an individual, but the substance of that individual within that district proposal could have been cut in the middle. In this way, the counterweight estimates that help them change the jump box of the district proposal.

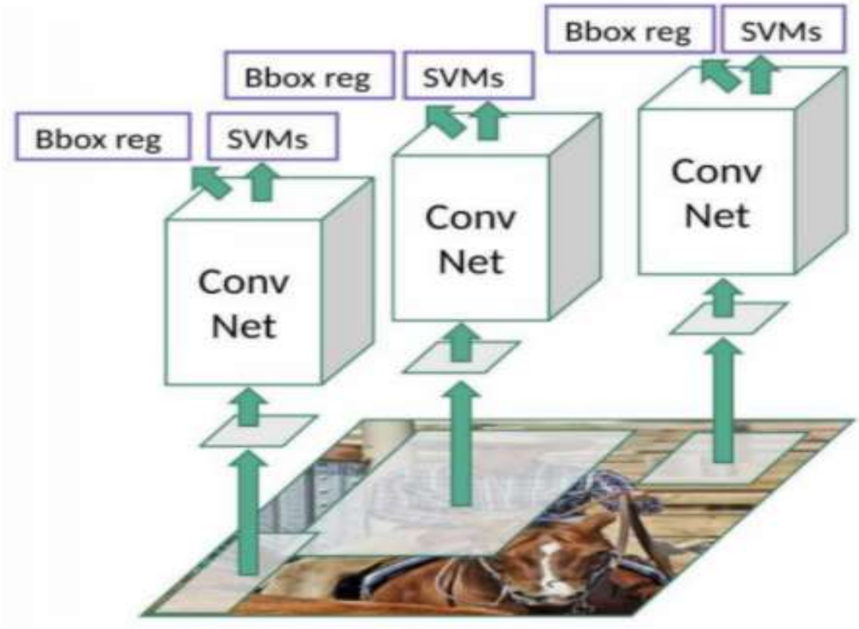


Figure 1-3 R-CNN Architecture (I)

Problems with R-CNN

It still takes a huge amount of time to train the network as you would have to classify 2000 region. It still sets aside an immense measure of effort to prepare the organization as you would need to arrange particle recommendations per picture. The specific pursuit calculation is a fixed. Subsequently, no learning is happening in this model.

1.4.1.2 Fast R-CNN

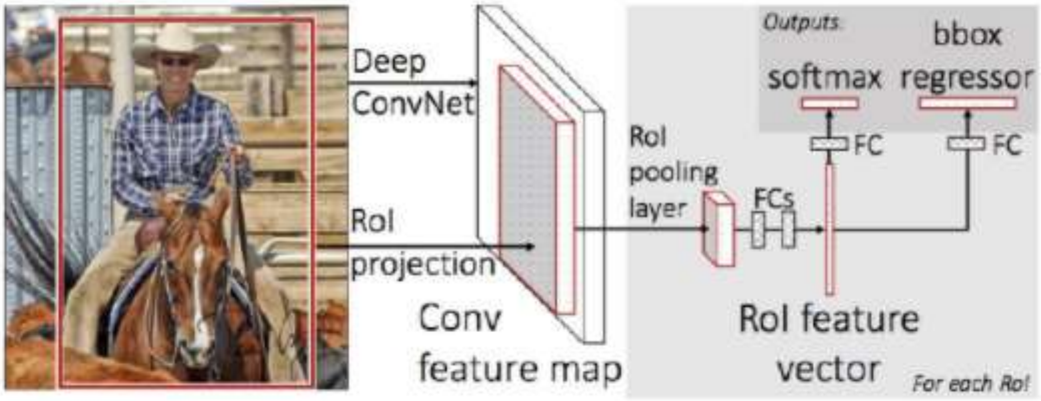


Figure 1-4 Fast R-CNN Architecture

A similar creator of the previous document (R-CNN) explained a portion of R-CNN's disadvantages to assemble a faster item location calculation and was called Fast R-CNN. The methodology is like the R-CNN calculation. However, instead of taking care of locale recommendations to CNN we feed the information image to CNN to produce a convolutional inclusion map. From the convolutional highlight map, we can recognize the proposition's locale and twist them into squares and by using a RoI grouping layer we reshape them to the fixed size so that it tends to be cared for in a fully associated layer. From the RoI include vector, we can use a softmax layer to anticipate the proposed locale class and also the counterweight estimates for the jump box. The "Quick R-CNN" explanation is faster than R-CNN is on the grounds that you don't need to take care of the 2000 district proposal to the convolutional neural organization without fail.

Given all things, convolution activity is constantly performed only once per image and an element map is created from it.

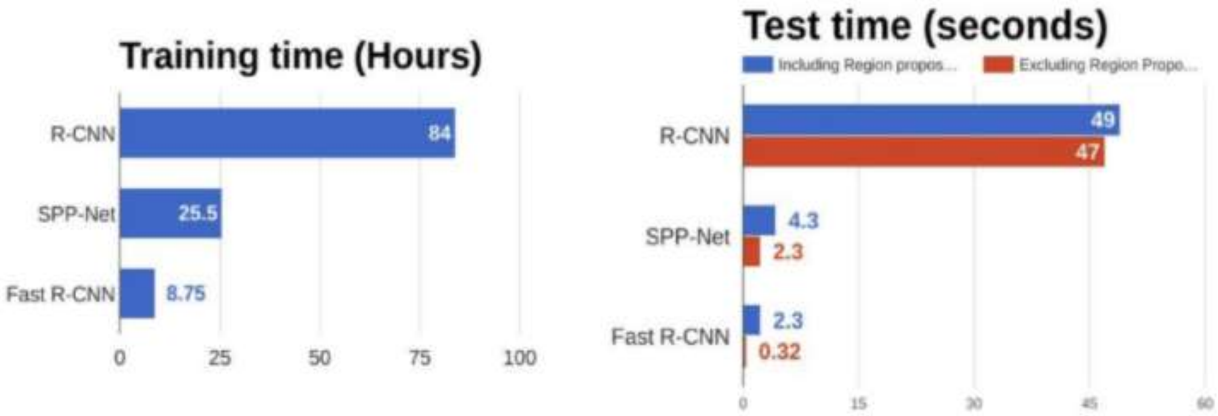


Figure 1-5 Fast R-CNN Training and Testing

From the above charts, you can construe that Fast R-CNN is essentially quicker in preparing and testing meetings over R-CNN. At the point when you take a gander at the exhibition of Fast R-CNN during testing, including district proposition make it difficult to calculate altogether when contrasted with not using area recommendations. Consequently, the locale which is proposition become bottlenecks in Fast R-CNN calculation influencing its presentation.

1.4.1.3 Faster R-CNN

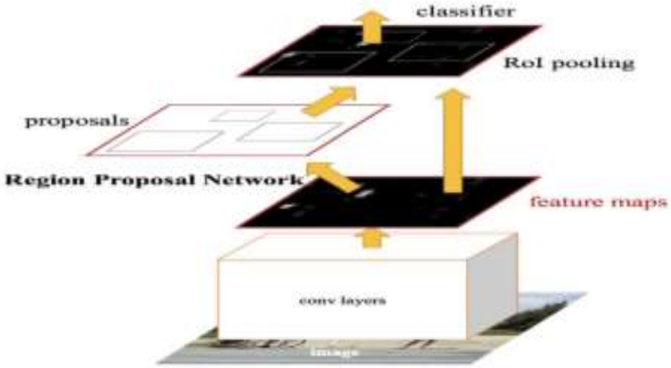


Figure 1-6 Faster R-CNN Architecture

Both previous calculations (R-CNN and Fast R-CNN) use a particular search to discover the locale proposal. Particular research is the moderate and tedious cycle that influences the presentation of the organization. Like Fast R-CNN, the image is given as a contribution to a convolutional network that gives a convolutional inclusion map. Instead of using the particular search calculation for the component guide to recognize the locale proposal, a different organization is used to anticipate area recommendations. The expected area that is recommendations are reshaped using a RoI grouping layer that is used to characterize the image within the proposed locale and forecast balance estimates for the bounce boxes.

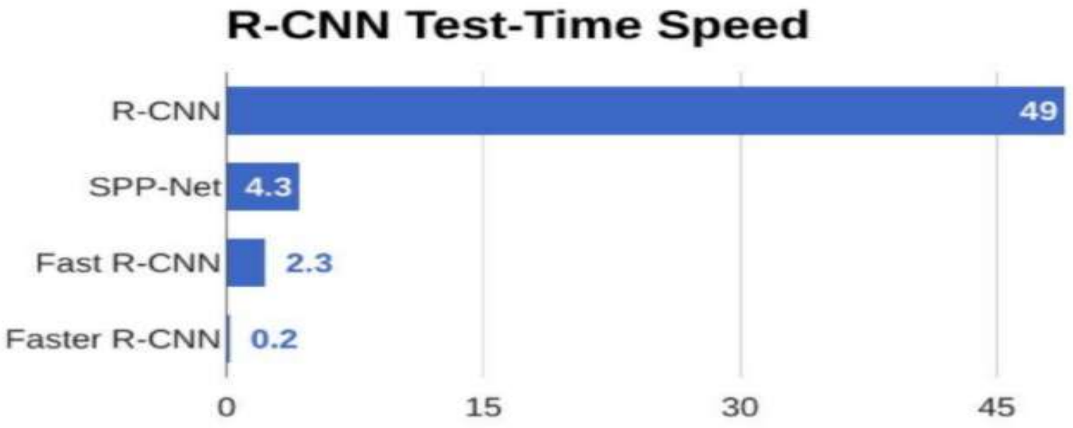


Figure 1-7 R-CNN Test-Time Speed

From above given diagram, it can be seen that Faster R-CNN is a lot quicker than its archetypes. Thusly, it can even be utilized for ongoing item discovery.

1.4.1.4 YOLO — You Only Look Once

All previous article discovery calculations have used districts to limit the element within the image. The organization doesn't take a look at the full photo. All things considered, parts of the image that are high probability of containing the article. YOLO or You Only Look Once is a much item recognition calculation that is not the same as the district-based calculations you saw earlier. At YOLO, a solitary convolutional network predicts jump boxes and class probabilities for these cases.

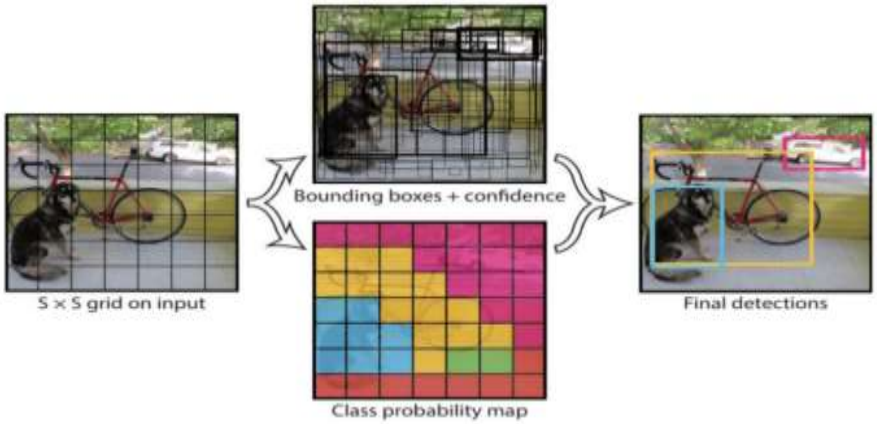


Figure 1-8 YOLO process

YOLO works by taking a photo and dividing it into an SxS matrix, inside each of the networks we take bounce boxes. For each of the jump boxes, the organization gives a performance a class probability and balance estimates for the bounce box. Jump boxes have the probability of class over an estimate of border being chosen and used to find the element within the image. YOLO is significantly grades faster (45 bushings for every second) than some other item recognition calculations. The impediment to YOLO's calculation is that it struggles with small items within the image, for example, you may experience problems distinguishing a herd of flying creatures. This is due to spatial limitations of the calculation.

1.4.1.5 SSD

Two main section for SDD model are:

1. Extract component maps, and

2. Apply convolution channels to identify objects.

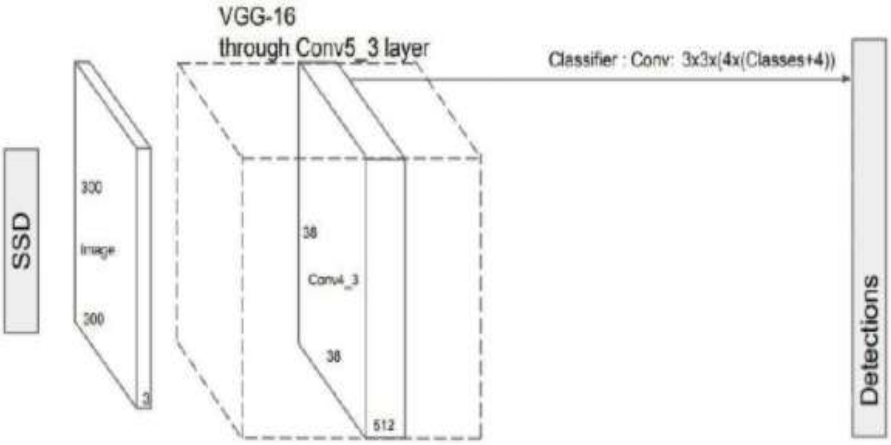


Figure 1-9 SSD Layers Diagram

SSD uses VGG16 to extend you to include maps. At that point it identifies the objects that use the Conv4_3. For delineation, we attract the Conv4_3 to be 8 × 8 spatially (must be 38 × 38). For every phone in the picture (also called area), it makes 4 expectations of elements.

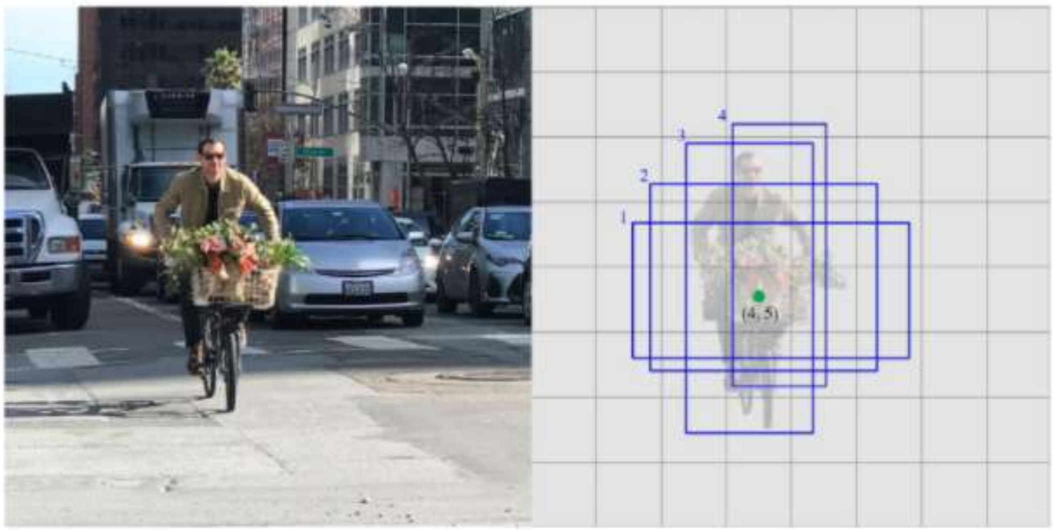


Figure 1-10 SSD detection

Each expectation comes out of a bounding box and 21 scores for each class (an additional class for no item), and we choose the most notable score as the class for the limited item. Conv4_3 makes absolute 38 × 38 × 4 forecasts: four expectations for each cell that places little importance on the depth of the featured maps. A normal, numerous forecasts do not contain any items. SSD saves a class "0" to display.

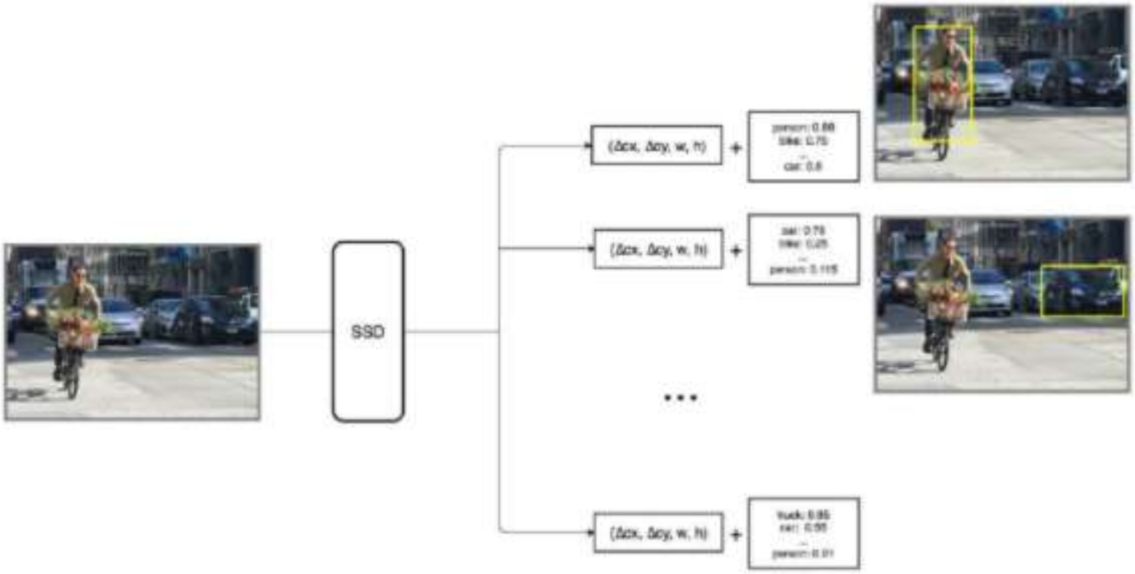


Figure 1-11 SSD working flow

SSD doesn't utilize the assigned district proposition organization. All things being equal, it makes plans to a simple methodology. Calculate area and class scores using small convolution channels. After extraction the component maps, SSD applies 6 convolution channels for each cell to make expectations. Each channel yields as 25 channels: 21 scores for each class plus a bounding box.

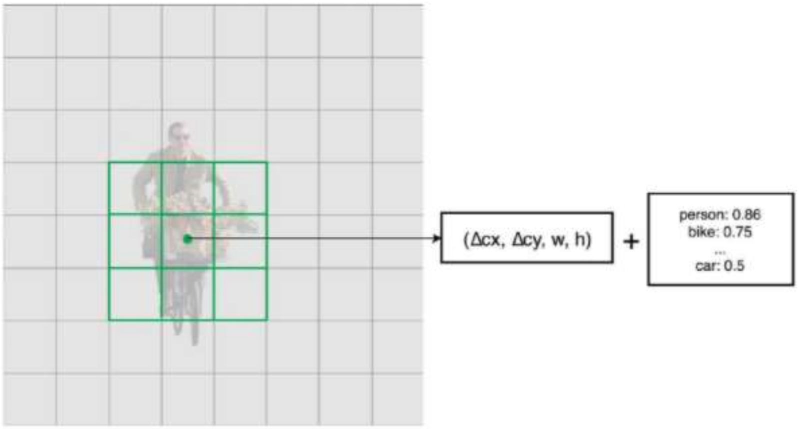
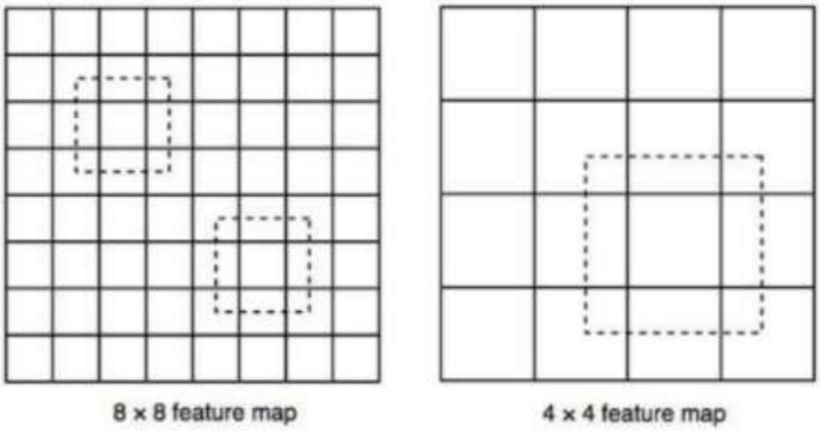


Figure 1-12 SSD with accuracy

Starting, we depict the SSD recognizes objects from a solitary layer. As a matter of fact, it utilizes various layers (multi-scale include maps) for the identifying objects autonomously. As CNN lessens the spatial measurement progressively, the goal of the component maps likewise decline. SSD utilizes lower goal layers for the identification of bigger scope objects. For instance, the 4× 4 component maps are utilized for the bigger scope object.



SSD adds 6 more helper convolution layers to picture after VGG16. Five of these layers will be added for object discovery. In which three of those layers, we make 6 expectations rather than 4.

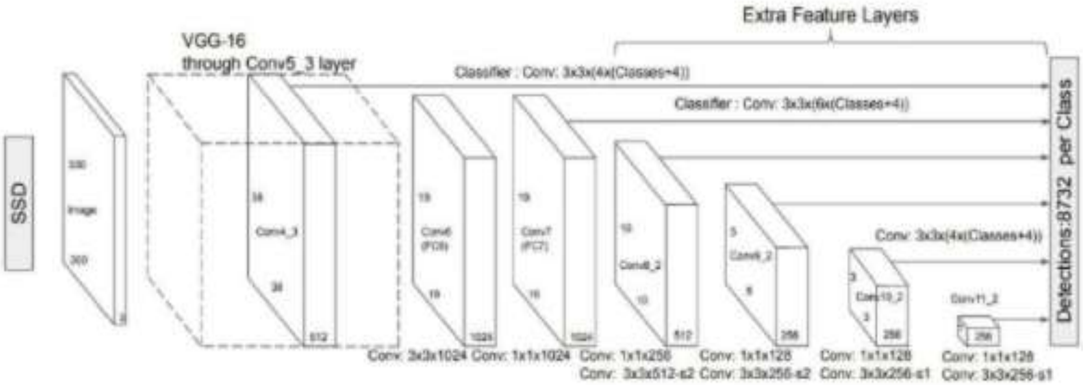
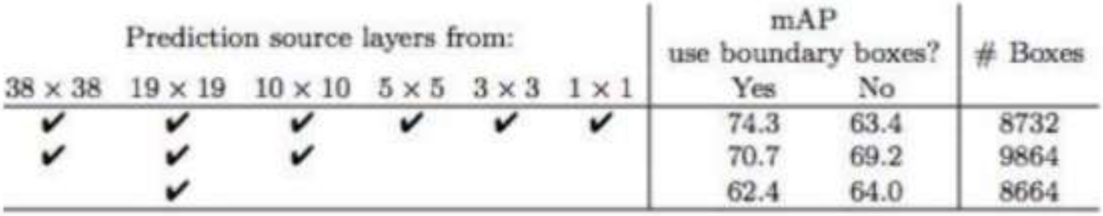


Figure 1-13 SSD Architecture

Multi-scale highlight maps upgrade precision. The precision with various number of highlight map layers is utilized for object recognition.



Prediction source layers from:						mAP use boundary boxes?		# Boxes
38×38	19×19	10×10	5×5	3×3	1×1	Yes	No	
✔	✔	✔	✔	✔	✔	74.3	63.4	8732
✔	✔	✔				70.7	69.2	9864
	✔					62.4	64.0	8664

Figure 1-14 SSD Prediction Matrix

1.4.1.6 MANet

To keep up ongoing velocities without giving up exactness in different item finders portrayed above, Liu et al proposed the SSD which is quicker than YOLO and has a practically identical precision to that of the most developed district based objective indicators. SD consolidates relapse thought of YOLO with the anchor box component of Faster R–CNN, predicts the item district dependent on the element guides of the diverse convolution layers, and yields discretized multiscale and multi corresponding default box arranges. The convolution part predicts outline facilitates pay of a progression of up-and-comer outlines and the certainty of every classification. The nearby component guides of multiscale zone are utilized to acquire results for each position in the whole picture. This keeps up the quick attributes of YOLO calculation and furthermore guarantees that the casing situating impact is like that is actuated by the Faster R–CNN. Be that as it may, SSD straightforwardly and freely utilizes two layers of the spine VGG16 and four additional layers got by a convolution with step 2 to develop highlight pyramid however needs solid logical associations. To fathom these problems, a single-stage recognition design, normally alluded to as MANet, which totals highlight data at various scales. MANet achievs 82.7% mAP on the PASCAL VOC 2007 test.

1.4.2 Color Models

Colors models give a standard method to determine a specific tone, by characterizing a 3D facilitate framework, and a subspace that contains all constructible tones inside a specific model. Each shading model is situated towards either explicit equipment (RGB, CMY, YIQ), or picture preparing applications (HSI).

1.4.2.1 The RGB Model

In the RGB model, an image contains three free image planes, one in all fundamental tones: red, green, and blue. To demonstrate a particular tone is to decide the proportion of all the basic portions present. The figure shows the mathematics of the RGB concealment model for demonstrating colors using a Cartesian stimulation structure. The grayscale range, such as hatches created using equal proportions of each fundamental, is located on the line that joins the high differentiation vertices.

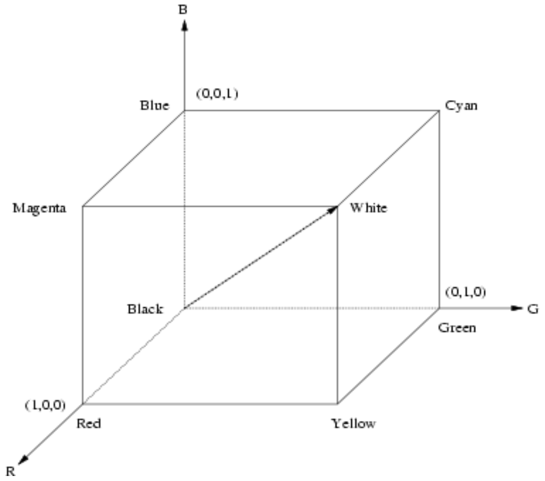


Figure 1-15 The RGB color cube. The greyscale spectrum lies on the line joining the black and white vertices.

1.4.2.2 The CMY Model

The CMY (cyan-magenta-yellow) model is a subtractive model suitable for tonemosis, for example in view of the tones in the paintings. While the RGB model asks what is added to boring to get a particular tone, the CMY model asks what is inferred from the target. For this circumstance, the primaries are cyan, garnet, and yellow, with red, green, and blue as discretionary shades. When a cyan-covered surface is illuminated by white light, no red light is reflected, and equivalently for fuchsia and green, and yellow and blue. The partnership between RGB and CMY models is given by:

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Figure 1-16 CMY Formula

1.4.2.3 The HSI Model

Shading might be determined by the three values of tone, immersion and force. It’s called HSI model, here is its representation.

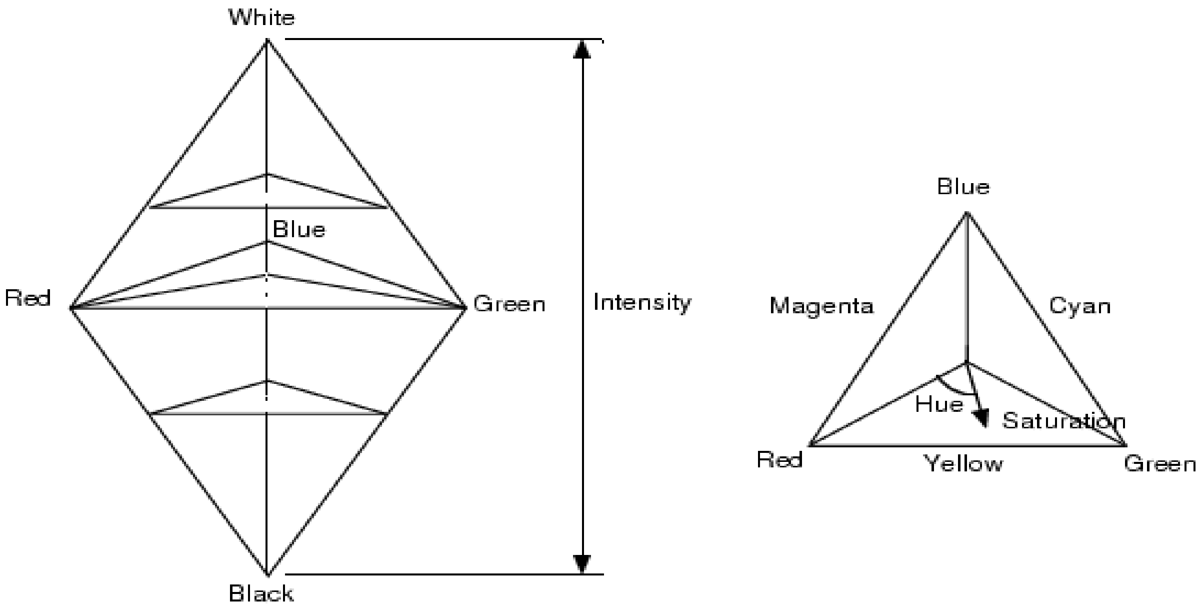


Figure 1-17The HSI model, showing the HSI solid on the left, and the HSI triangle on the right,

Shaped by taking a horizontal cut through the HSI strong at a specific power. Shadings on the outside of the strong are completely immersed, for example unadulterated shadings, and the greyscale range is on the hub of the strong. The intensity is given by

$$I = \frac{R+G+B}{3},$$

As R, G and B are components of these colors, between the range of [0, 1]. The intensity represent the average of these colors components.

$$S = 1 - \frac{\min\ (R,G,B)}{I} = 1 - \frac{3}{R+G+B} \min\ (R,G,B)$$

If the R, G or B having minimum values or equal to zero than the color will be pure one.

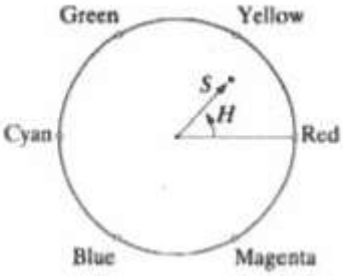


Figure 1-18. HSI Model

The biggest application of this model is that it represents colors similarly to how the human eye senses of colors.

1.5 Analysis from Literature Review

A task is accomplished with the specified goal of achieving a desired output, it is necessary to go through past research and techniques to evaluate the details. Different approaches have been used in past few decades for object detection. These techniques differ from each other, bring to light the merits of each and limitations with respect to our problem. We use mobile net model for object detection in our project to train it over CIFAR-100 dataset.Color detection is the essential step in many computer vision systems. In this project, Python is used to detect a specific color in a given image. Image processing in Python provides a variety of reference algorithms, methods, and applications for image manipulation, visualization, and segmentation. An image is an array of pixel values. The image can be represented with many color models such as grayscale, RGB, HSV, etc.

Here a shading histogram strategy is utilized for color classification which depends on a specific shading space, for example, RGB or HSV. The shading Histogram of a picture speaks to the conveyance of the shading arrangement in a picture. It shows various sorts of tones showed up and the quantity of pixels showed up in each kind of shading.

1.5.1 Mobile Net

Mobile Net is a convolutional neural network design model for image classification and mobile vision. There are several models, in addition, which makes Mobile Net exceptional than less estimation ability to execute or apply motion classification in some way to. This makes it an ideal fit for mobile gadgets, implanted frames and PC without GPU or low computational capacity with compromise through and through the accuracy of the results.

1.5.1.1 MobileNet Architecture

- MobileNets is invented for smart devices and embedded systems, which used depth wise separable convolutions to build deep light weight neural networks based on a streamlined architecture.

- Two simple global hyper parameters are introduced that efficiently compensate between inertia and accuracy

The central Layer of MobileNet model are de separable channels of depth, called prudent deeply separable convolution. Organizational structure is another factor in supporting exposure. Finally, the width and lens can be adjusted for the balance between inertia and accuracy.

1.5.1.2 Depth wise Separable Convolution

Definable convolutions of depth that is a form of factored convolutions that factor a standard convolution into a wise deep convolution and a convolution 1×11×1 called wise convolution point. In MobileNet, deep prudent convolution applies a single filter to each input channel. The wise point convolution then applies a convolution 1×11×1 to combine the outputs of the convolution wise depth. The following figure illustrates the difference between standard convolution and wise depth separable convolution.

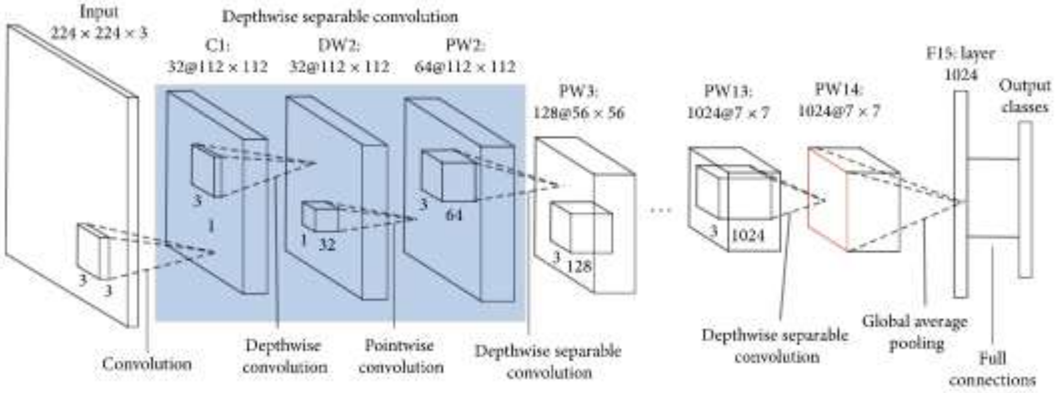
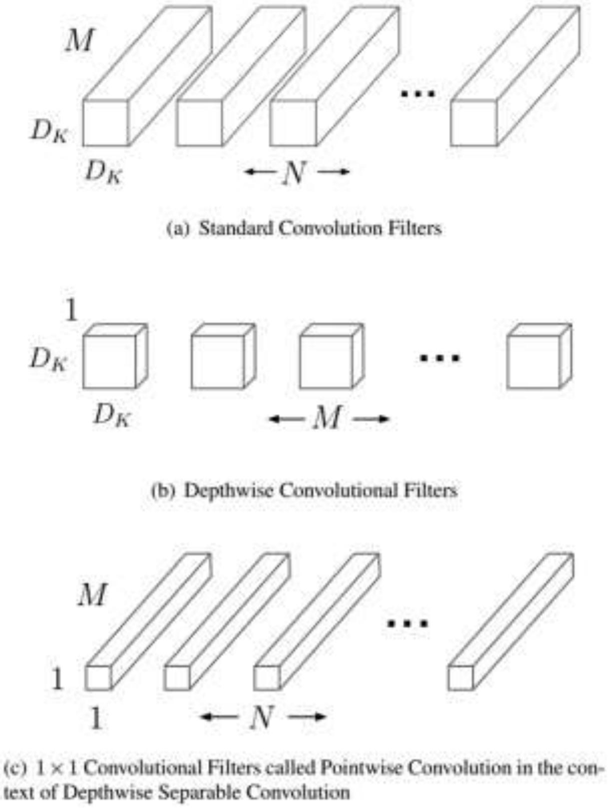


Figure 1-19 MobileNet model



(a) Standard Convolution Filters

(b) Depthwise Convolutional Filters

(c) 1 × 1 Convolutional Filters called Pointwise Convolution in the context of Depthwise Separable Convolution

Figure 1-20 MobileNet Layers

The standard convolution has cost

D_k · D_k · M · N · D_F · D_F

Depth wise separable convolution cost is

D_k · D_k · M · D_F · D_F + M · N · D_F · D_F

1.6 Methodology and Software Lifecycle for this Project

Project methodology is an important phase of any project because it is a key element and set the overall tone. For this we must first understand the steps that are involved in project methodology. We use iterative approach to develop this application.

1.6.1 Methodology

Software development methodology involves splitting the work into different stages with the goal of better management, planning and understanding Software.

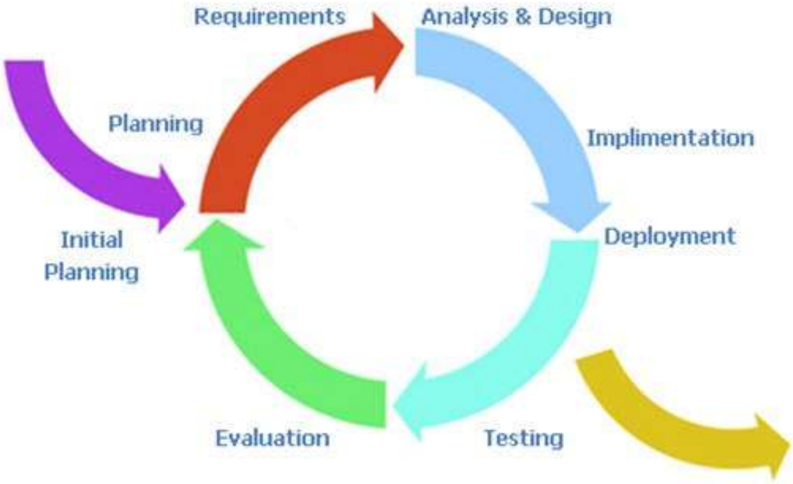


Figure 1-21 Iterative Model

1.6.2 Rationale behind selected methodology

Iterative process is the way to breakdown the whole process to develop a complete software in one go into smaller chunks. The reasons behind this approach is that it is easy measurable and small chunks are easy to test and debug as compare to developing the whole software and then testing it, at this point debugging becomes difficult.

Chapter 2

PROBLEM DEFINITION

2 Problem Definition

Multiple external devices and mobile applications used for object or color detection. Low vision user unable to read text labelling on screen. The main problem if low vision and color blind user wants to recognize an object than how he/she will do it? By downloading multiple android apps for multiple purpose which is not the effective way and resources and a lot of effort required.

2.1 Problem Statement

We have seen many projects but they are all performing a single task. If someone is hired to detect objects, they will download the Object Detectors app. And if someone is hired to detect color, they will download the color Detectors app. Different external devices also present for different purpose. What if a person is suffering from multiple disabilities? These all devices and multiple application is not effective for such type of person because it required more efforts and extra money. We have created this app to get rid of these problem. This could be accomplished by providing low vision and color-blind people with information about the things which will be helpful for them. That's why we built this project which perform more task at a time.

2.2 Deliverables and Development Requirements

List of deliverables of our project:

- Smart Phone
- Image/Camera View

Development requirements include following software requirements:

- Android Studio & SDK (Designing and Development)
- Python
- Anaconda

- Programming language : Python, Java ,XML
- Firebase (Database)

Chapter 3

REQUIREMENT ANALYSIS

3 Requirement Analysis

Software Requirement Analysis (SRS) is provides the basic understanding of functional as well as non -functional requirements. We can consider it as a starting point of project because it serves as written contract for client about the features and functionalities of the project. With the help of SRS user can easily understood.

3.1 Use Case Diagram

A sensible image of the associations among the segments of a structure is called use case outline. A use case diagram is a technique utilized as a part of framework investigation to perceive and arrange framework necessities. We design UML (Unified Modelling Language) figures to model a framework in the simple and efficient way. UML describes several types of the diagram to cover all aspects of the framework because it is not enough to define all aspects of the system using single UML diagram.

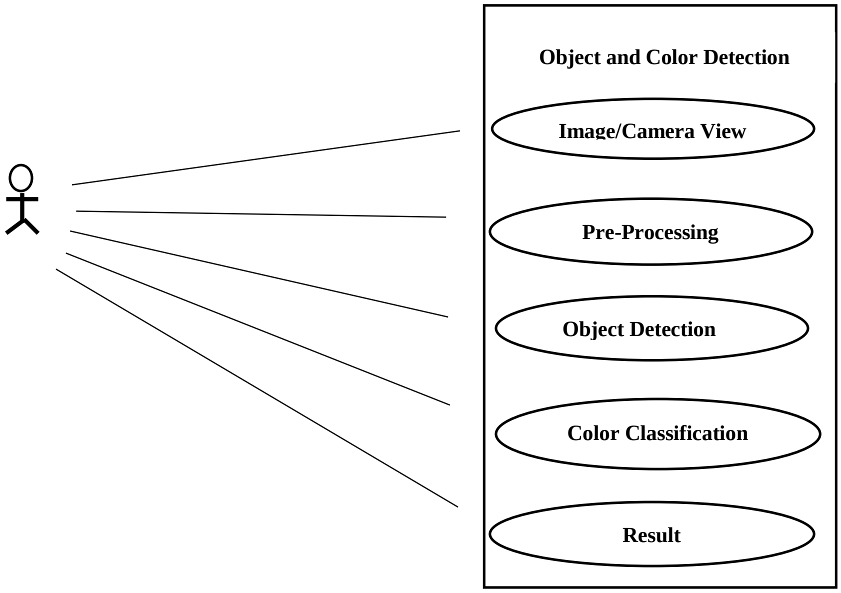


Figure 3-1 Use Case Diagram

3.1.1 Detailed Use Case Diagram

Use case figure is one of the UML diagrams which models dynamic behavior. It is dynamic in nature, consisting of some inside or outside components for making the connection. It is a set of use cases, actors (primary and secondary actor) and relationships. A use case diagram is used to represent a particular functionality of a system. Many use cases diagrams are used to model the entire system. It is a generic way to define that use case diagram captures the all dynamic aspects of the entire system.

1. Use-case name: Object & Color Image /Camera View.

Action: User.

Description: User will input an image from system directory to application or live view as initial step to use this application.

Pre-condition: User must provide Object and Color image to a system as an input and it must be in .PNG format.

Post-condition: After giving an input to a system, image is uploaded.

2. **Use-case name**: Pre-Processing.

Action: Computer-aided system.

Description: An input image is going for pre-processing.

Pre-condition: Image must be uploaded.

Post-condition: After giving an input to a system, image is pre-processed, and user can exit or it go for object detection process.

3. Use-case name: Object Detection

Action: Computer-aided system.

Description: A pre-processed image is sent forward for Object detection.

Pre-condition: Input pre-processed images or new image is uploaded.

Post-condition: After object detection, object name label will be shown on screen.

4. Use-case name: Color Classification.

Action: Computer-aided system.

Description: An image is sent forward for color classification.

Pre-condition: Input images or new image is uploaded.

Post-condition: After classification, color are extracted.

5. Use-case name: Result (Labeled and audio)

Action: Computer-aided system

Description: After classification of each Object and Color class, each Object and Color label is classified with promising accuracies and yield output.

Pre-condition: The labels must classify correctly so the system play it in audio.

Post-condition: Result is displayed after correct classification of Object and Color app along with their labels and accuracies.

3.2 Functional Requirement

Functional requirement describes the predictable behavior of the framework. These are main framework requirement and without applying any of these requirements the framework should be fragmented.

- The system will take the input image and perform pre-processing.
- The system will detect object in the image and classify the color.
- Based on the classification, the system will show output on the screen and in audio.

3.2.1 Input image

Identifier	FR-1
Title	Input image /Camera View
Requirements	The system take image
Dependencies	None

Table 3-1 Input Image

3.2.2 Pre-processing

We convert the images from .mat format into .jpg and label them. We make separate directories of each type of class. And then we apply other elements of preprocessing like, image scaling, color space transformation, contrast enhancement, image restoration and noise removal.

3.2.3 Object Detection

Object detection will be carried out using machine learning model. Once the user is movement and the surrounding object is detected the specific item is matched with object saved inside the database.

3.2.4 Color Classification

Color will be classify by using histogram method which based on intensity of color in pixels of image.

3.2.5 Results/Output

Identifier	FR-3
Title	Output
Requirements	After Object and Color detection display label on screen with accuracy and play audio.
Dependencies	FR-2

Table 3-2 Result

3.3 Non-Functional Requirement

3.3.1 Performance

Performance of our system is efficient, and it takes very less time to perform the action. Our CAD system saves the time.

3.3.2 Capacity

Our system must meet the agreed capacity.

3.3.3 Availability

The system will be available into any hospital or clinic.

3.3.4 Reliability

Our system is 100% reliable and generate an accurate result.

3.3.5 Maintainability

With the time and needs our system will be maintained properly and any feedback from the user is our priority while performing maintenance.

3.3.6 Efficiency

This system is efficient as it does not require any kind of effort to use it and also does not take much time.

3.3.7 Flexibility

The system provides the user to load the image easily.

3.3.8 Usability

This system is very easy to use by radiologist and physicians. The designed interface is very easy to understand.

Chapter 4

DESIGN AND ARCHITECTURE

4 Design and Architecture

After gathering all requirements, the next step is to start planning that how we are going to develop our project, how much resources, costs, time, benefits and other items are required. After planning we move to the designing and architecture phase that which techniques and methods we can use and how we are going to develop our project. It is the most challenging phase of project development.

4.1 System Architecture

We have used CNN based mobile net model in which we used image and label from dataset to train it and after object detection we use color histogram classification for the classification of color in give image or camera view. We start with pre-processing, extracting images, object detection classify color, and show output on screen and play in Audio also. Each step is discussed in detail.

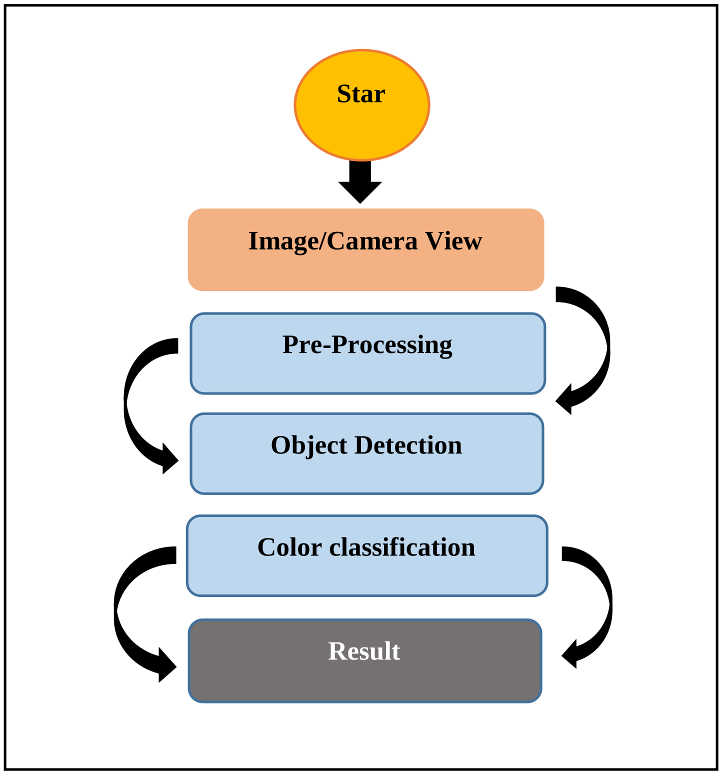


Figure 4-1 System Architecture

4.2 Data Representation [Diagram + Description]

A data flow diagram is graphical interpretation of information move from a data framework is called data flow diagram. A DFD is used for basic step to create an overview of the system without going into great aspect, which can later be elaborated.

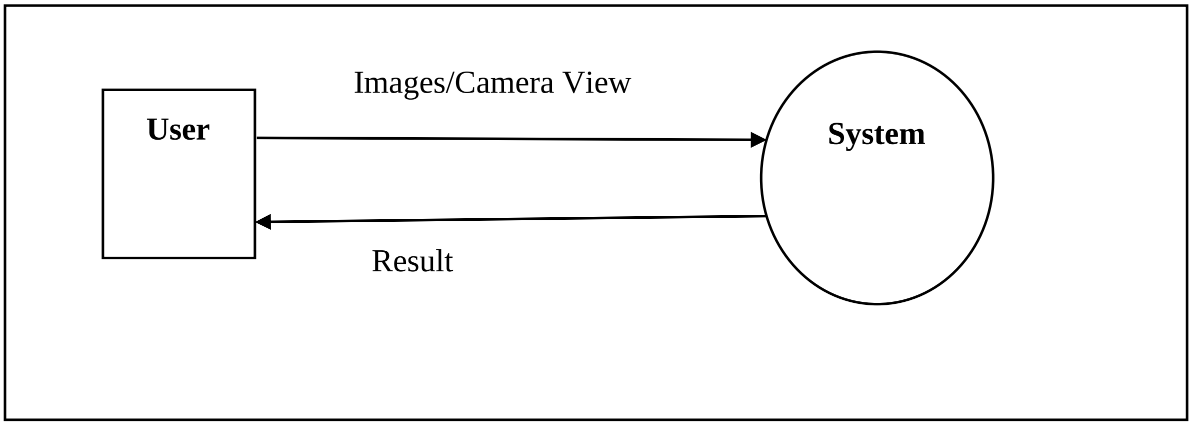


Figure 4-2 Data flow diagram

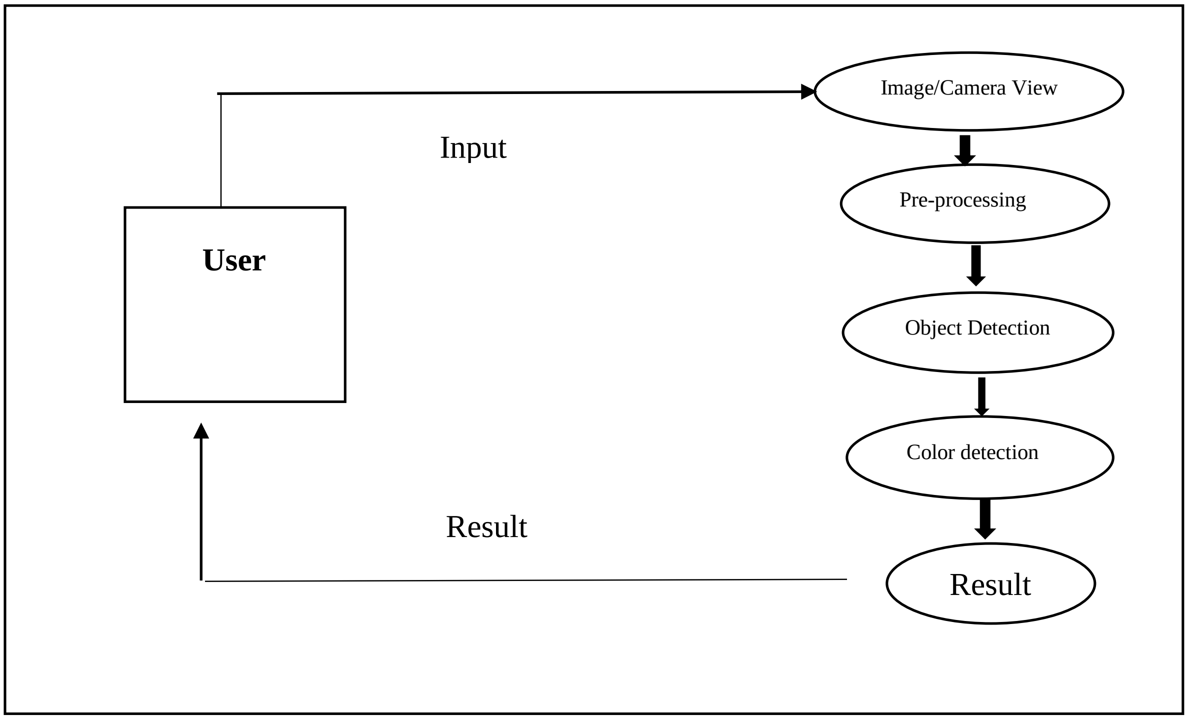


Figure 4-3 Data flow diagram detail

4.3 Process Flow / Representation

4.3.1 Activity Diagram

An activity diagram is depicting the dynamic component of the framework. An activity diagram is essentially a flowchart to speak from one action to another movement. The movement may be relating to an operation of the framework. This diagram is used to explain the dynamic features of the system. It is more like a flow chart because it shows the flow of data from one activity to other.

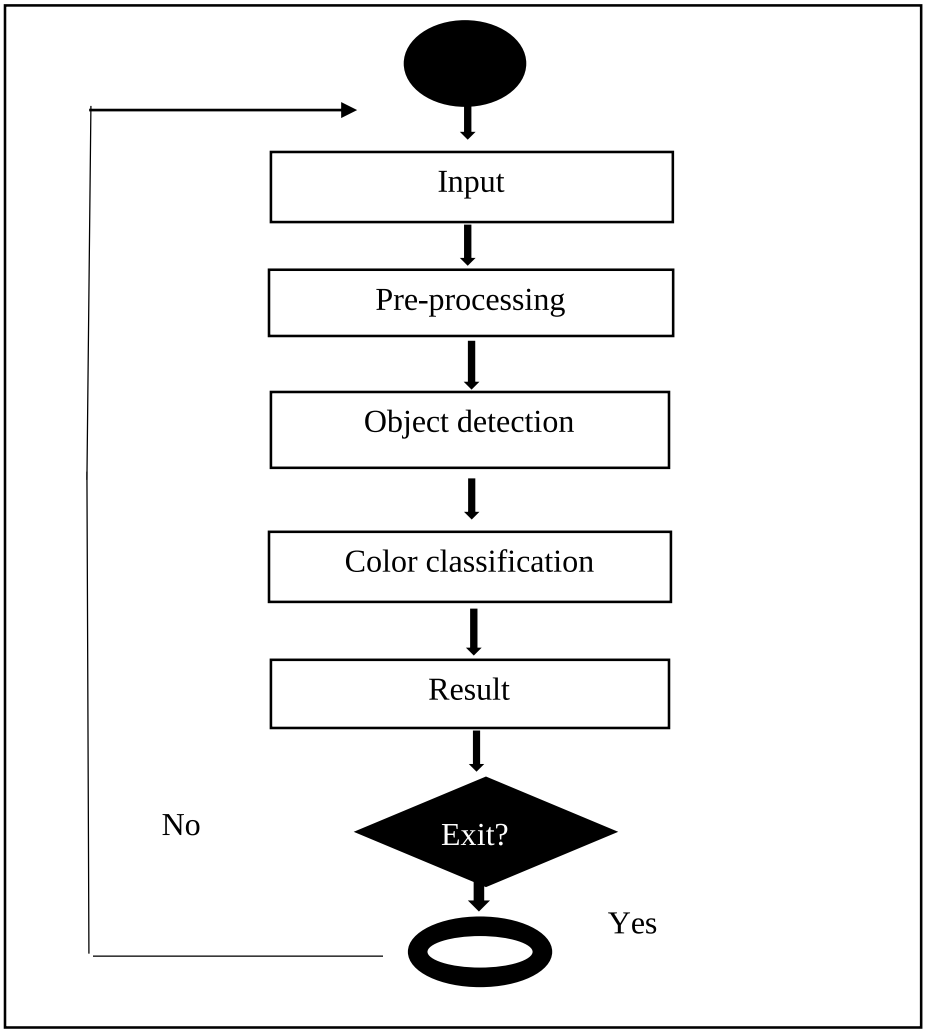


Figure 4-4 Activity Diagram

4.4 Design Models

4.4.1 Sequence Diagram

A cooperation graph that shows how objects work with each other and in what request is called sequence diagram. It is built up of message arrangement outline.

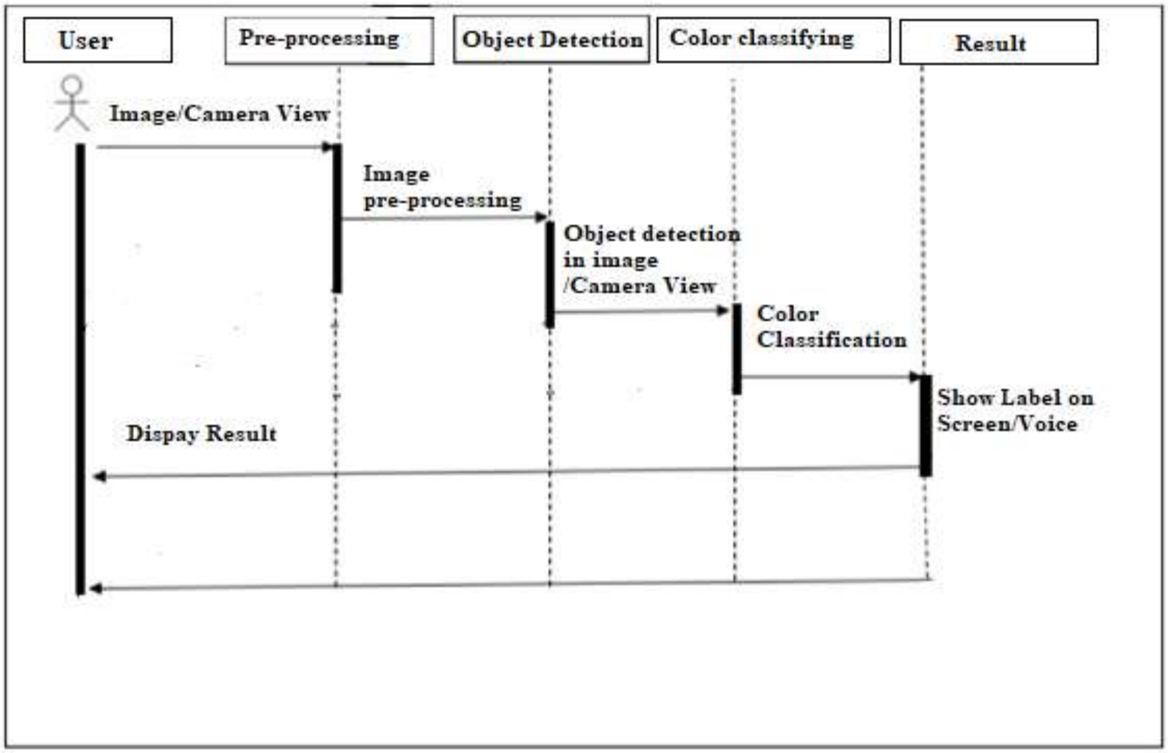


Figure 4-5 Sequence Diagram

5 References

1. https://www.researchgate.net/publication/337464355_OBJECT_DETECTION_AND_IDENTIFICATION_A_Project_Report
2. https://medium.com/analytics-vidhya/image-classification-using-mobilenet-in-the-browser-b69f2f57abf
3. https://www.cs.toronto.edu/~kriz/cifar.html
4. https://link.springer.com/chapter/10.1007/978-3-030-49076-8_9
5. https://yourstory.com/mystory/mobile-apps-important-business
6. https://www.sciencedirect.com/topics/computer-science/android-application#:~:text=Android%20is%20an%20open%2Dsource,customize%20to%20differentiate%20their%20products.
7. https://www.techopedia.com/definition/25099/android-app
8. https://medium.com/mindorks/detection-on-android-using-tensorflow-a3f6fe423349
9. https://arxiv.org/abs/1704.04861
10. https://proceedings.neurips.cc/paper/2018/hash/9908279ebbf1f9b250ba689db6a0222b-Abstract.html
11. https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53
12. https://jivp-eurasipjournals.springeropen.com/articles/10.1186/s13640-018-0258-x
13. https://www.sitepoint.com/a-short-guide-to-color-models/
14. https://www.cronj.com/blog/color-space-image-processing/#:~:text=In%20Digital%20Image%20Processing%2C%20the,with%20colors%20as%20humans%20interpret.
15. https://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/OWENS/LECT14/lecture12.html
16. https://en.wikipedia.org/wiki/Color_histogram
17. https://en.wikipedia.org/wiki/Color_histogram#:~:text=In%20general%2C%20a%20color%20histogram,such%20as%20RGB%20or%20HSV.&text=This%20process%20is%20called%20color,color%20histogram%20of%20the%20image.