

Virtual Assistant with Embedded Image Recognition

A PROJECT REPORT

Submitted By

BIBIN BENNY (Reg. No : PJR17CS008)

JABIR MOHAMMED (Reg. No : PJR17CS013)

KAVYA BABU (Reg. No : PJR17CS015)

KRISHNAPRIYA S NAIR (Reg. No : PJR17CS016)

to

The APJ Abdul Kalam Technological University

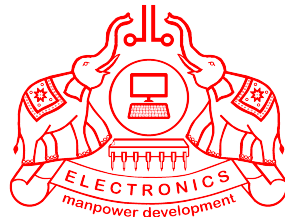
in partial fulfillment of the requirements for the award of the Degree

of

Bachelor of Technology

in

Computer Science and Engineering



Department of Computer Science and Engineering

College of Engineering Poonjar

Poonjar Thekkekkara P O, Kerala, India 686582

JUNE 2021

DECLARATION

I undersigned hereby declare that the project report “ Virtual Assistant with Embedded Image Recognition ”, submitted for partial fulfillment of the requirements for the award of degree of Bachelor of Technology of the APJ Abdul Kalam Technological University, Kerala is a bonafide work done by me under supervision of Mr.Anychal J Vattakunnel, Assistant Professor, Dept. of Computer Science & Engineering, College of Engineering Poonjar. This submission represents my ideas in my own words and where ideas or words of others have been included, I have adequately and accurately cited and referenced the original sources. I also declare that I have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title of any other University.

Poonjar
21/06/2021

BIBIN BENNY
JABIR MOHAMMED
KAVYA BABU
KRISHNAPRIYA S NAIR

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
COLLEGE OF ENGINEERING POONJAR



CERTIFICATE

This is to certify that the report entitled **Virtual Assistant with Embedded Image Recognition** submitted by **Mr.Bibin Benny** to the APJ Abdul Kalam Technological University in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineering is a bonafide record of the project work carried out by him under our guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

Mr.Anchal J Vattakunnel
Project guide

Mr.Rajesh K R
Project Coordinator

Mr. Rajesh K R
HOD

CONTENTS

ACKNOWLEDGMENT	i
ABSTRACT	ii
LIST OF FIGURES	iii
1 INTRODUCTION	1
2 LITERATURE SURVEY	3
3 SYSTEM DESIGN	4
3.1 DISADVANTAGES OF THE EXISTING SYSTEM	4
3.2 PROPOSED SYSTEM	4
3.2.1 Object Detection	5
3.2.2 Voice Detection Model	6
4 DESIGN	7
5 IMPLEMENTATION	8
6 RESULTS	9
7 CONCLUSION	11
8 SCREENSHOT	12
REFERENCES	22

ACKNOWLEDGEMENT

It is a great pleasure to acknowledge all those who have assisted and supported me for successfully completing my project.

First of all, I thank God Almighty for his blessings as it is only through his grace that I was able to complete my project successfully.

I take this opportunity to extend my sincere thanks to my project guide Mr. Anchal J Vattankunnel, Assistant Professor, Dept. of Computer Science, College of Engineering Poonjar for his constant support and immense contribution for the success of my project.

I also extend my sincere thanks to all other members of the Department of Computer Science & Engineering for sharing their valuable comments during the preparation of the project.

I am also grateful to Prof. Rajesh K. R., Project Coordinator and Head of Computer Science & Engineering Department, for the valuable guidance as well as timely advice which helped me a lot during the preparation of the project.

I am deeply indebted to Dr. Jobymol Jacob, Principal, College of Engineering Poonjar for her encouragement and support.

I whole - heartedly thank all my classmates, for their valuable suggestions and for the spirit of healthy competition that existed between us.

ABSTRACT

In the upcoming age of digital era, the virtual assistants are booming business, with specific skill sets, they can serve multiple clients and work on different projects. They work according to the client and offer a wide range of affordable services. With their top-notch assistance, you'll get things done easily within the deadline. The proposed is a virtual assistant with embedded image recognition capabilities to assist the end user through Weather forecasting, Launching applications, News imparting, Real time image Recognition. Interaction can be done through a microphone or through written commands. The object detection component allows the users to identify the thing around him/her with ease, and that too on a low spec device. The assistant uses NLP along with Tensorflow to perceive the texts and understand what the user wants. Yolo algorithm along with Tensorflow Object Detection API and Opencv are used to specifically identify the images.

LIST OF FIGURES

3.1	Architecture	5
3.2	Overview of Object Detection	6
3.3	Voice Detection Cloud API	6
6.1	Real-Time Object Detection on COCO	9
6.2	Scaled-YOLOv4: Scaling Cross Stage Partial Network	10
6.3	Metrics for Yolo-V4 without CUDACNN Support	10
8.1	fig1	12
8.2	fig2	13
8.3	fig3	13
8.4	fig4	14
8.5	fig5	15
8.6	fig6	16
8.7	fig7	17
8.8	fig8	18
8.9	fig9	19
8.10	fig10	20
8.11	fig11	21

CHAPTER 1

INTRODUCTION

In the upcoming age of digital era, the virtual assistants are booming business, with specific skill sets, they can serve multiple clients and work on different projects. They work according to the client and offer a wide range of affordable services. With their top-notch assistance, you'll get things done easily within the deadline.

Intelligent Personal Assistant Automated Personal Assistant or Automated Virtual Personal Assistant can perform tasks, or services, on behalf of an individual based on a combination of user input and location awareness. It has the ability to access information from a variety of online sources such as weather conditions, traffic congestion, news, stock prices, user schedules, and retail prices. Many CEOs and business executives need to hire personal assistants to help them with various essential tasks. Different responsibilities such as dealing with correspondences, replying to emails, managing diaries, scheduling appointments, organizing meetings, and booking events among others require an entirely dedicated personal assistant to carry them out. Business executives and owners need to be aware of specific information such as directions, retail prices, weather conditions, traffic news of specific locations and other kinds of information that impact their businesses to make informed decisions. Customers also need personal assistants to answer their queries and to help them with various basic tasks and company information such as products availability, how long it will take to receive their orders and more. Business owners also need to send email newsletters or update their social page status in a specific time when they are more likely to get huge responses and to convert more of their target audience. Intelligent personal assistants have been engineered to help with these basic tasks.

Best Intelligent Personal Assistants or Automated Personal Assistants: Google Assistant, Nina, Viv, Jibo. Google now. Hey Athena, Cortana, Mycroft. Braina Virtual Assistant Siri, SILVIA. Amazon Echo, Bixby. Lucida. Cubic Dragon Go, Hound, Aido, Ubi Kit BlackBerry

Assistant. Maluuba. Vlingo are some of Top Intelligent Personal Assistants or Automated Personal Assistants.

The features provided by these assistant are :

- **Calendar Meeting Reminder:** Intelligent Personal Assistant schedule meetings appointments reminders instantly on behalf of the user. IDAs also helps the user remember everything they have set it to remind them and send the user signals, photos, links and more via SMS, emails or other means. This software can set alarms to tell the user of an upcoming event or task
- **Automation:** Help to automate most of the essential functions that the user wants. The user can utilize IPAs to do research, identify landmarks, shop. and translate foreign languages among other tasks.
- **Natural Conversation:** Intelligent Personal Assistant can understand and respond to complex questions. It recognizes the intent of the user's inquiry. personalizes the responses based on context, and troubleshoots the problem using conversational strategies when answering social questions, reacting to customer frustrations and even becomes a live chat agent when need be. The user can create reminders, ask questions and even type anything they want by speaking to the program
- **Smarter Learning.** Artificial intelligence technology uses both machine learning and natural language understanding which allows it to obtain industry- specific knowledge and unique business data and thus can do marketing for an enterprise.

However with all these features they are bound to the internet and cannot process without a significant network speed. Also considering that most of these don't support real time object detection they somewhat fail to serve their purpose as virtual assistant.

In such a Constance our project aims to adhere an answerable virtual assistant along with real time image recognition capabilities.

The proposed is a virtual assistant with embedded image recognition capabilities(Buddy) to assist the end user through Weather forecasting, Launching applications, News imparting, Real time image Recognition. Interaction can be done through a microphone or through written commands. The object detection component allows the users to identify the thing around him/her with ease, and that too on a low spec device. The assistant uses NLP along with Tensorflow to perceive the texts and understand what the user wants. Yolo algorithm along with Tensorflow Object Detection API and Opencv are used to specifically identify the images.

CHAPTER 2

LITERATURE SURVEY

The era of virtual assistant was brought around 2003 with DARPA. The application was then sought by Apple in 2010, but the service was removed from the Apple store. In 2010, Apple invented its own virtual assistant Siri. Which at that point was highly impactful to the virtual assistant enthusiasts.

Siri faces tight competition from its competitor Google Assistant that interacts with the user using its Dialogflow API, which is based on NLP and ML. The query is then passed to the cloud engine and a significant response is computed.

Google Assistant allows you to basically enquire about anything, be it questions about the weather, flight status, or places; Google Assistant does the search for you. On top of that, you can have information translated in over 100 languages. Google Home lets you get help around the house through the hands-free speech recognition capability to manage your schedule and control your smart home. It allows you to direct the assistant to do things like; making a reservation, setting a reminder, or playing your favorite playlist, all this directly from your phone. Android Wear helps you make use of Google Assistant.

Alexa is Amazon's cloud-based voice service available on hundreds of millions of devices from Amazon and third-party device manufacturers. With Alexa, you can build natural voice experiences that offer customers a more intuitive way to interact with the technology they use every day.

Cortana is an intelligent personal assistant created by Microsoft that offers personal digital assistance through reminders and helping you work across devices.

CHAPTER 3

SYSTEM DESIGN

3.1 DISADVANTAGES OF THE EXISTING SYSTEM

The Existing Model is equipped with the following features

- Weather forecasting.
- Launching applications.
- News imparting.
- Knowledge Sources.
- Multidevice Connectivity.
- TTS(Text to Speech) and STT(Speech to Text)

However with all these features they are bound to the internet and cannot process without a significant network speed. Also considering that most of these don't support real time object detection they somewhat fail to serve their purpose as virtual assistant.

3.2 PROPOSED SYSTEM

The Proposed is equipped with the following features in-addition to the features of the existing model

- Character Detection.
- Language Translation.
- Real Time Image Recognition.

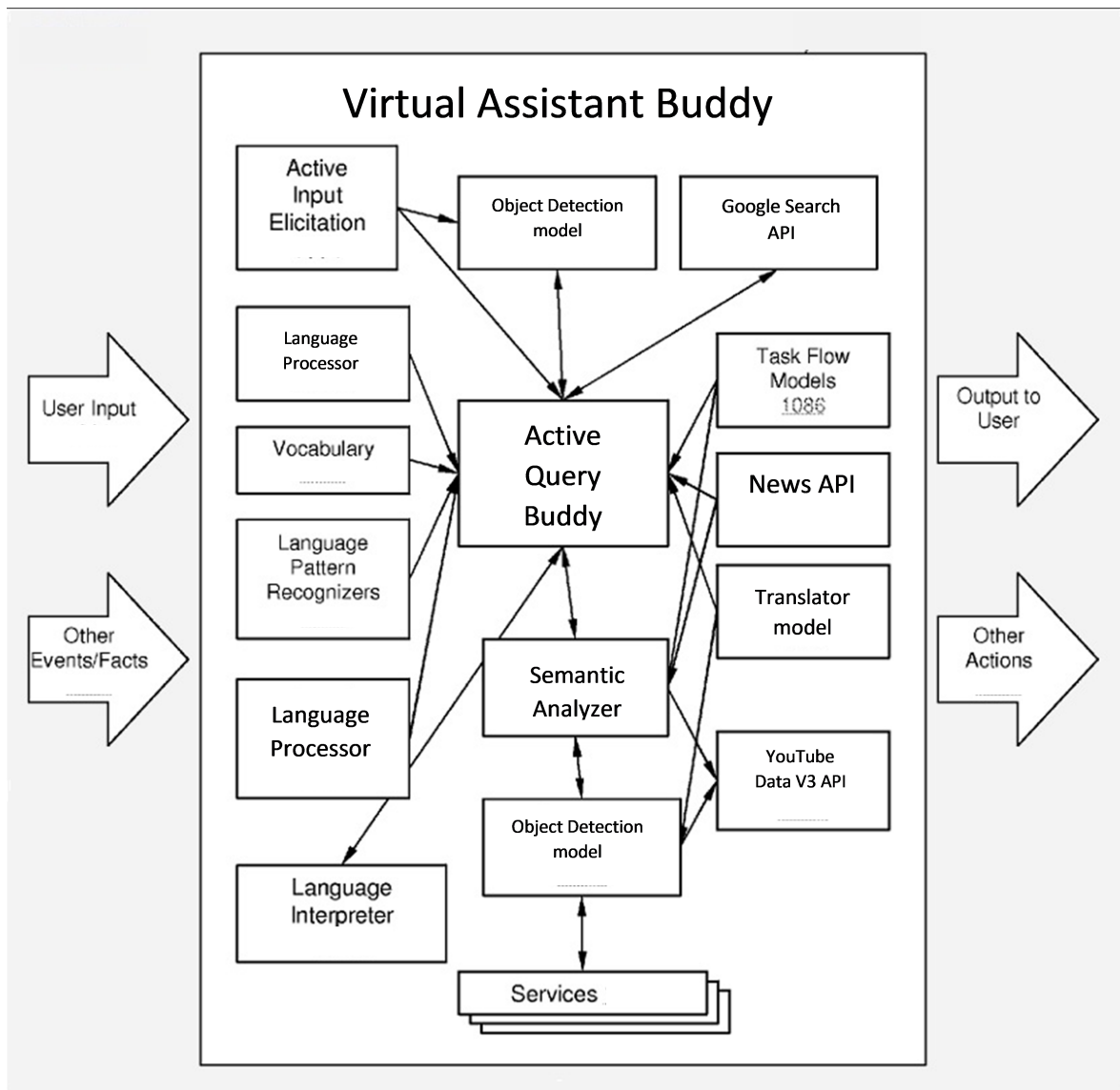


Figure 3.1: Architecture

3.2.1 Object Detection

Object recognition is a general term to describe a collection of related computer vision tasks that involve identifying objects in digital photographs.

Image classification involves predicting the class of one object in an image. Object localization refers to identifying the location of one or more objects in an image and drawing abounding box around their extent. Object detection combines these two tasks and localizes and classifies one or more objects in an image.

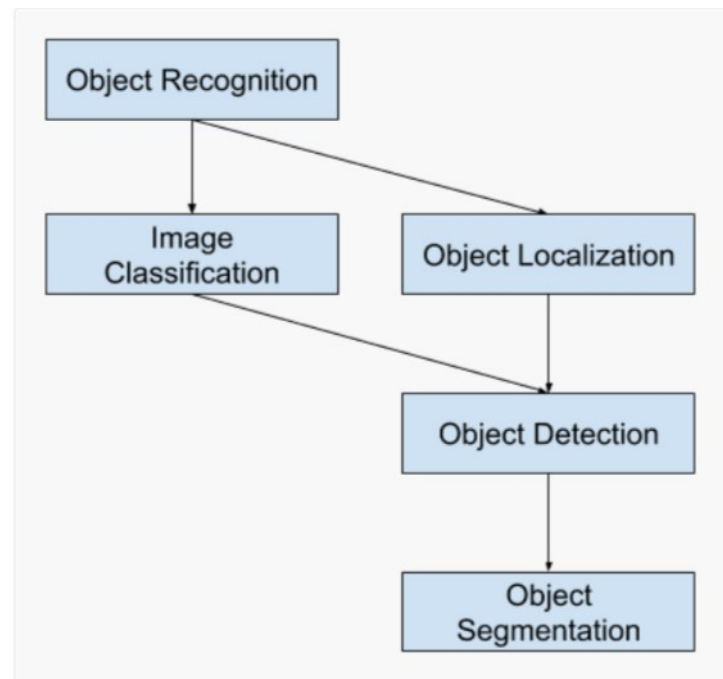


Figure 3.2: Overview of Object Detection

3.2.2 Voice Detection Model

The Alan Cloud is the default deployment scenario. In this scenario, all resources such as VMs with voice scripts and VUI speech components are hosted in the public cloud. Alan owns and manages these resources and provisions them to you based on your needs.

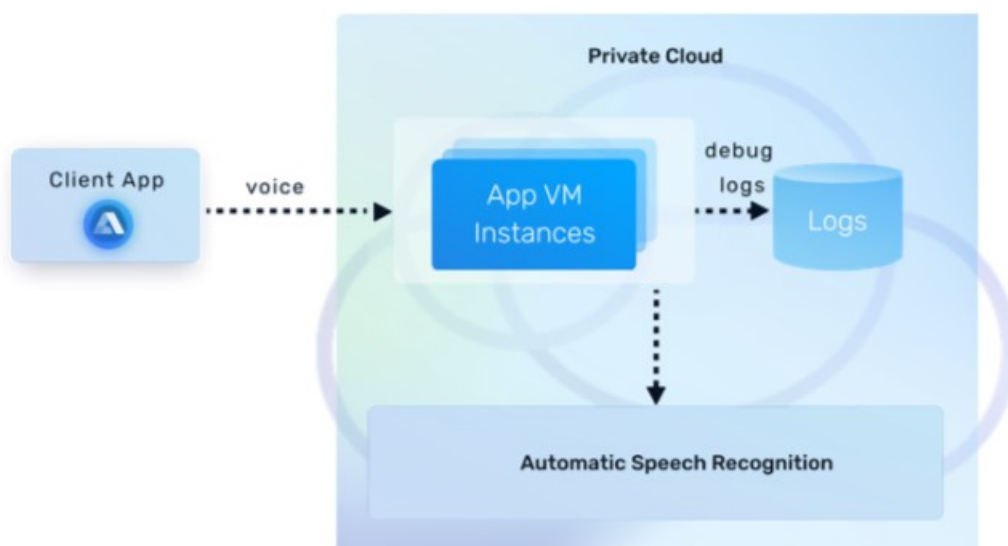


Figure 3.3: Voice Detection Cloud API

CHAPTER 4

DESIGN

The Virtual assistant mainly provides a platform for the user to interact with the assistant either via voice or textual commands.

The working of the application is as follows:

- The VA accepts input as voice which takes the query and interprets the suitable command.
- After the successful completion of a query request from the user the VA then focuses its attention to other commands. These mainly include the news impartment, Speech recognition, text translation, etc .
- The VA instructs the user to act accordingly to the commands such that the commands are successfully executed and the user intention is satisfied.
- The object detection is done with the help of Tensorflow library. Tensorflow bundles together Machine Learning and Deep Learning models and algorithms. Tensorflow allows developers to create a graph of computations to perform.

CHAPTER 5

IMPLEMENTATION

- OBJECT DETECTION

- You only look once (YOLO) is a state-of-the-art, real-time object detection system. On a Pascal Titan X it processes images at 30 FPS and has a mAP of 57.9
- YOLOv3 is extremely fast and accurate. In mAP measured at .5 IOU YOLOv3 is on par with Focal Loss but about 4x faster. Moreover, you can easily tradeoff between speed and accuracy simply by changing the size of the model, no retraining required!
- Prior detection systems repurpose classifiers or localizers to perform detection. They apply the model to an image at multiple locations and scales. High scoring regions of the image are considered detections.
- We use a totally different approach. We apply a single neural network to the full image. This network divides the image into regions and predicts bounding boxes and probabilities for each region. These bounding boxes are weighted by the predicted probabilities.

- TRAINING MODEL

- COCO is a large-scale object detection, segmentation, and captioning dataset. This version contains images, bounding boxes, labels, and captions from COCO 2014, split into the subsets defined by Karpathy and Li (2015). This effectively divides the original COCO 2014 validation data into new 5000-image validation and test sets, plus a "restval" set containing the remaining 30k images. All splits have caption annotations.

CHAPTER 6

RESULTS

The proposed virtual assistant with embedded image recognition capabilities(Buddy) successfully assists the end user through Weather forecasting,Launching applications, News imparting,Real time image Recognition.The object detection component satisfyingly allows the users to identify the thing around him/her with ease,and that too on a low spec device.The assistant uses NLP along with Tensorflow to perceive the texts and understand what the user wants.Yolo algorithm along with Tensorflow Object Detection API and Opencv are used to specifically identify the images.

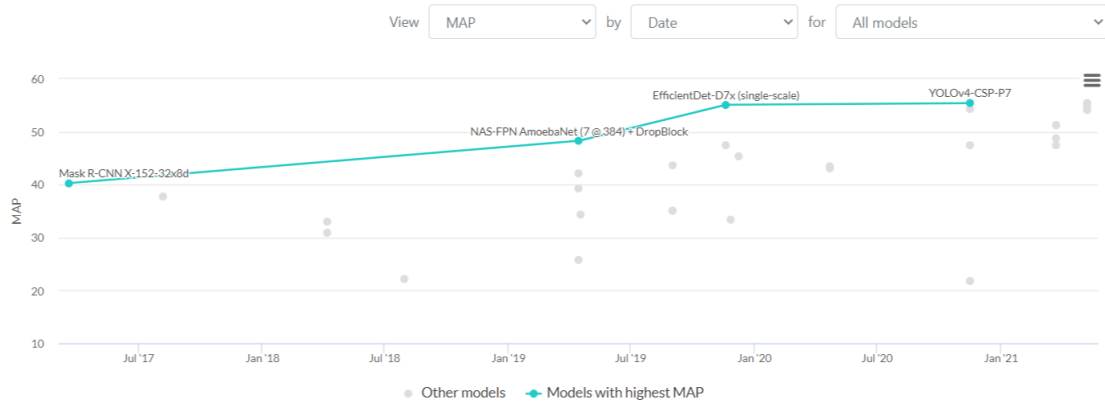


Figure 6.1: Real-Time Object Detection on COCO

The MS COCO (Microsoft Common Objects in Context) dataset is a large-scale object detection, segmentation, key-point detection, and captioning dataset. The dataset consists of 328K images.

Splits: The first version of MS COCO dataset was released in 2014. It contains 164K images split into training (83K), validation (41K) and test (41K) sets. In 2015 additional test set of 81K images was released, including all the previous test images and 40K new images.

Based on community feedback, in 2017 the training/validation split was changed from 83K/41K to 118K/5K. The new split uses the same images and annotations. The 2017 test set is a subset of 41K images of the 2015 test set. Additionally, the 2017 release contains a new unannotated dataset of 123K images.

The YOLOv4-tiny model achieves 22.0



Figure 6.2: Scaled-YOLOv4: Scaling Cross Stage Partial Network

Even without the TensorflowRT and CUDACNN Support, the model does not fail to show its efficiency

TensorRT,	FPSAGX	FPSNX	FPSX2	FPSNANO
without	120	75	42	16
with	290	118	100	39

Figure 6.3: Metrics for Yolo-V4 without CUDACNN Support

CHAPTER 7

CONCLUSION

An initial outcome of the study indicates that the Virtual Assistant are capable of helping the user in his/her end to end activities.

Intelligent virtual assistant increases company revenue by reducing the efforts in customer handling. It also reduces live chat volume, offers quick response, and saves time.

By enabling real time detection enables these to unleash the true hidden potential by helping the user to understand different languages and also proves as a boon to the blind.

In the rising age of automation and digitalization virtual assistants will play a crucial role

CHAPTER 8

SCREENSHOT

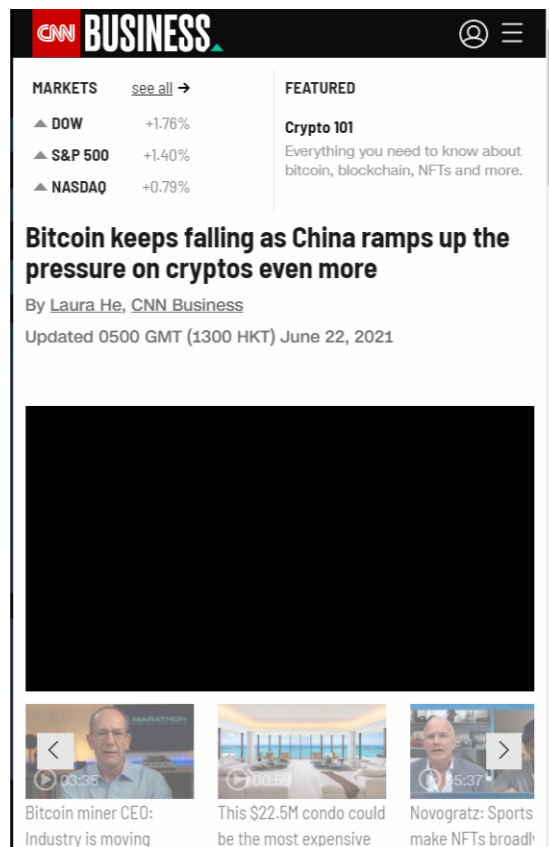


Figure 8.1: fig1

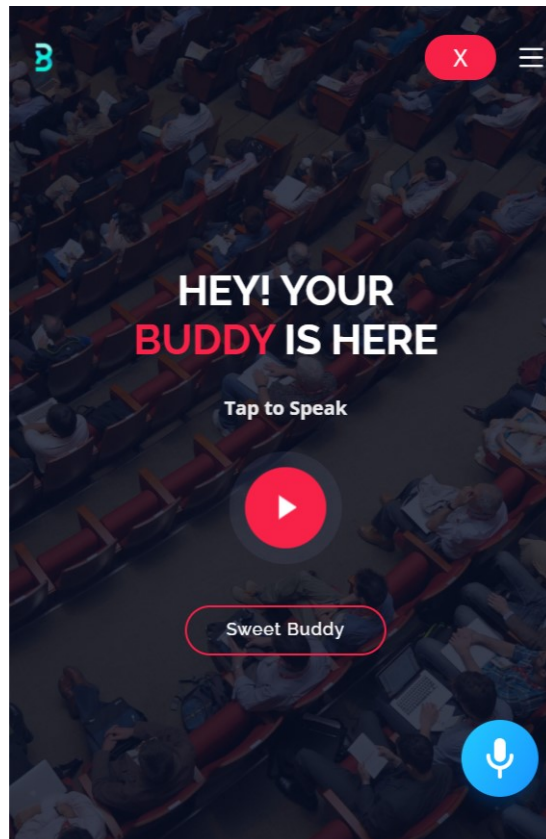


Figure 8.2: fig2

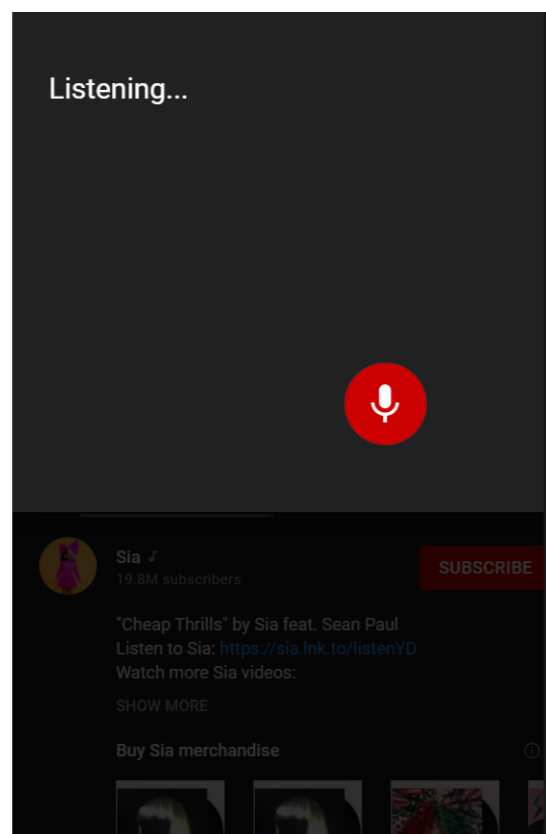


Figure 8.3: fig3

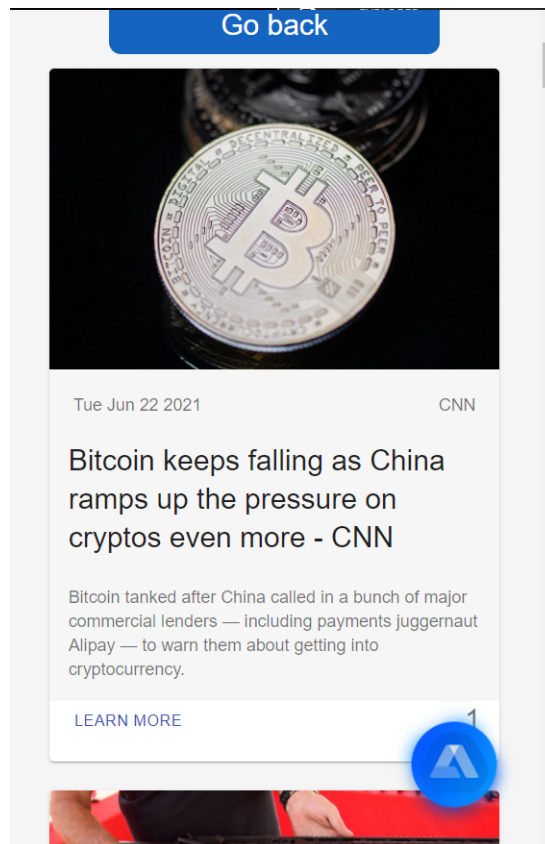


Figure 8.4: fig4

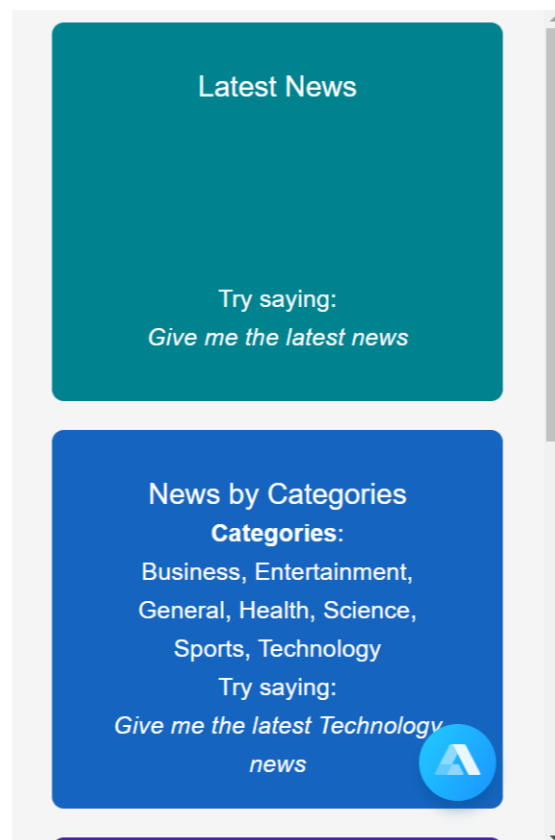


Figure 8.5: fig5

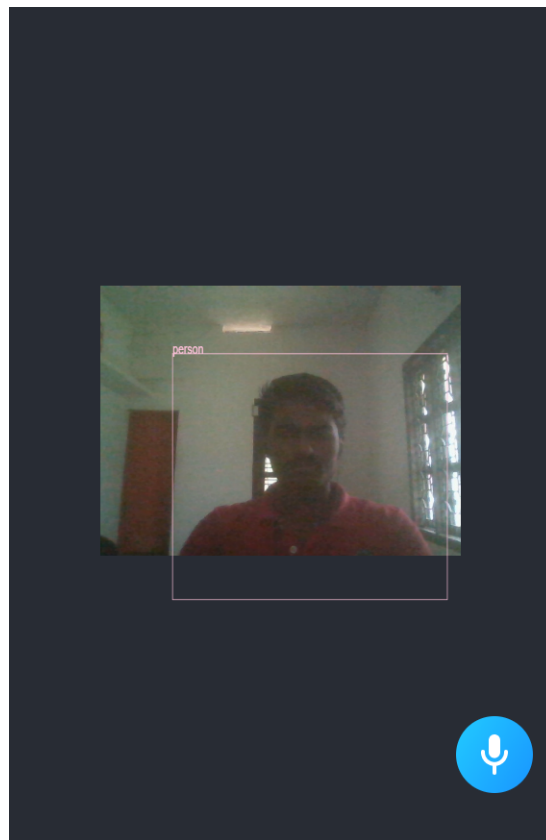


Figure 8.6: fig6

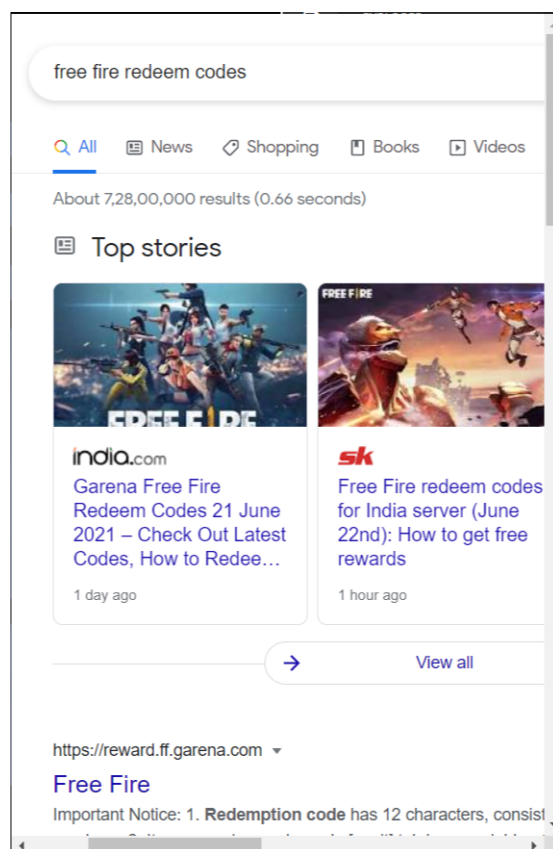


Figure 8.7: fig7

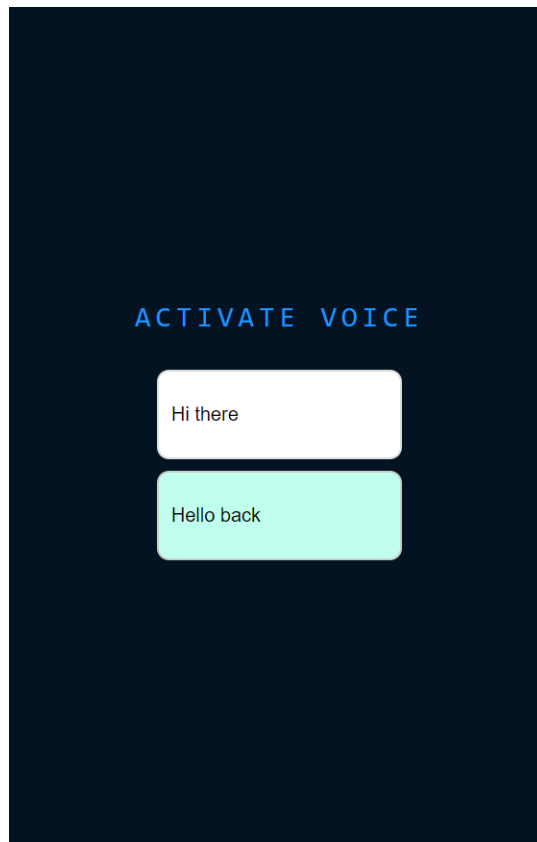


Figure 8.8: fig8

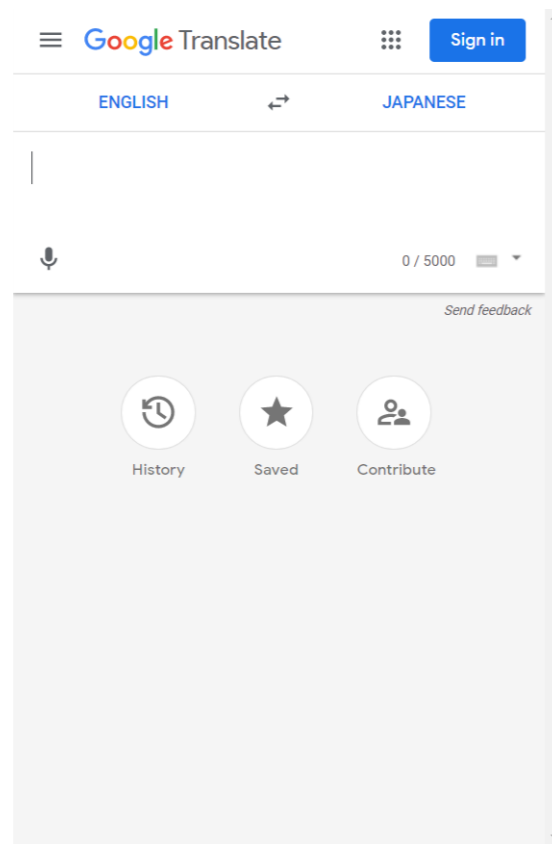


Figure 8.9: fig9

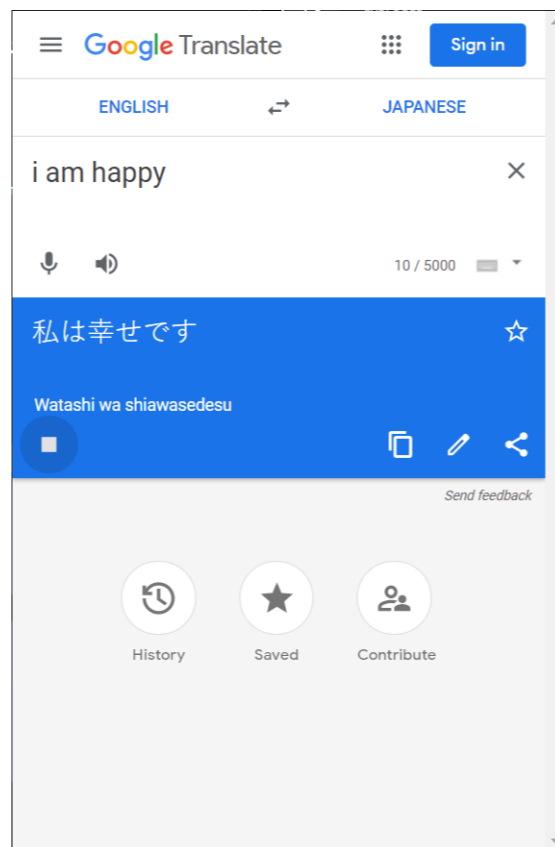


Figure 8.10: fig10



Figure 8.11: fig11

REFERENCES

- [1] <https://pjreddie.com/darknet/yolo/>
- [2] <https://machinelearningmastery.com/object-recognition-with-deep-learning/>
- [3] <https://paperswithcode.com/sota/real-time-object-detection-on-coco>
- [4] <https://alan.app/docs/usage/how-works/deployment>
- [5] <https://www.youtube.com/watch?v=2RxHQoiDctI>
- [6] <https://www.youtube.com/channel/UCHXa4OpASJEwrHrLeIzw7Yg>