A

PROJECT REPORT

ON

Hand-Sign Detection

In partial

fulfillment of

B. Tech IV year
Computer Science & Engineering

Session 2020-2021

Submitted To:

Mr. Rajendra Purohit
Head Of Department
Computer Science & Engineering

Submitted by:

Vaibhav Saran Rashi Singhvi
18EJICS169 18EJICS136
VII Semester

Department of Computer Science & Engineering
Jodhpur Institute of Engineering & Technology
Jodhpur 2018-2022

Certificate

This is to certify that the project entitled **"Hand Sign Detection"** has been carried out by the students of "**Jodhpur Institute of Engineering and Technology, Jodhpur"** under my guidance and supervision in partial fulfillment of the degree of Bachelor of Engineering in Computer Science & Engineering of Bikaner Technical University, Bikaner during the academic year 2021-2022.

Vaibhav Saran and Rashi Singhvi
Date: 23/10/2021
CSE Department
JIET Jodhpur

Ms. Arshi Riyaz
Assistant Professor(Sr. Scale)

Mr Rajendra Purohit
Head of Department, CSE

Acknowledgment

I would like to express my sincere thanks to **Mr. Nicholas Renotte**, for his work on Object Detection was a great help for me in making my own project and bringing it to a level where it can be used in the real world.

As this project is based on open source projects, I would like to express my gratitude to all open source developers and contributors for sharing their experiences and ideas.

I would also like to extend my warm gratitude towards my prestigious Institute: **Jodhpur Institute of Engineering and Technology** and **Mr. Rajendra Purohit** (Head of Department CSE) for providing me the opportunity to explore and build real-time working projects which increases my knowledge in the field of Deep Learning with Tensorflow.

I would like to express my special thanks to **Ms. Arshi Riyaz** for approving our project proposal and we assure you to fulfill your requirements from our side; also heartfelt gratitude from the team for guiding us throughout the project and sharing your ideas and opinion regarding our project.

Consent Letter From Client

This letter is regarding the proposal sent by you. I am glad to inform you that your proposal has been accepted and approved. From the project I expect the following deliverables:

- ➢ Real-time detections using a webcam or mobile camera
- ➢ 20 phrases and alphabets should be accurately translated
- ➢ A lightweight application that can easily be deployed on various platforms
- ➢ Proper interface

Other Requirements:

- ➢ Proper Documentation
- ➢ Verified license on Github, released as open-source
- ➢ Minimal Human Interaction
- ➢ Well performing Model

I am willing to act as a legitimate client for the project mentioned below.

Project Name: Hand Sign Detection

Date: 16 August 2021

Team Members: Vaibhav Saran and Rashi Singhvi

Client Details

Signature:

Name: Ms. Arshi Riyaz

Contact Number: +91 88904 18618

Address: JIET Jodhpur

Abstract

Hand gesture recognition is very significant for human-computer interaction. In this work, we present a novel real-time method for hand gesture recognition. In our project, the hand region is extracted from the background with the background subtraction method. Then, the palm and fingers are segmented to detect and recognize the fingers. Finally, a TensorFlow object detection technique is applied to predict the labels of hand gestures.

The model is designed by using SSD MobileNet V2 FPNLite 320x320, this will compress the image to 320x320 in the pre-processing, and in post-processing, it is going to take the detections that it found and convert it back to the original resolution. It uses Image augmentation, i.e. it might darken, shift, or flip the image so that we can ideally get a better performing model.

For Data Collection we will not use any images from third-party sources or any other applications. We have generated our own data by capturing pictures of ourselves from different angles to get better and accurate results. Then generate the respective XML file for each image and label them accordingly. There are 20 labels and 26 English alphabets used in this project. Per class 60 images will be collected which makes 2760 images, these images will be split into train and test datasets. The split ratio will be 80-20.

Table of Contents

1. Introduction

Motivation: Mute people often come across scenarios where they face a huge gap while communicating with normal people if a translator is not available.

Our project is basically designed to solve these kinds of real-life problems faced by common and disabled people. To fill the communication gap between them, we introduce our model that will detect hand-signs gestures and predict outcomes.

Objective: To use SSD MobileNet V2 FPNLite 320x320, an object detection model available as a part of Tensorflow model zoo, to make a lightweight model that will predict hand signs from live video.

Outcome: A lightweight model whose weights can be deployed on a website or any other application which can further be used to communicate with mute or deaf people.

Problem Statement: There are some state-of-the-art models available which have taken an approach to this problem by using 3D CNN and LSTM with FSM Context-Aware Model and many more. The general concept is that several CNN layers are used followed by several LSTM layers, use of pre-trained mobile net followed by many LSTM layers. These models end up requiring a large amount of data to produce good results and also demand very high computing power due to the presence of 30 to 40 million parameters. As a part of this project, we would like to work on a lightweight model which can produce the same results and consumes much less memory as well as data compared to SOTA architectures.

2. Literature Survey

Machine Learning[1]: Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy. Machine learning is an important component of the growing field of data science. Through the use of statistical methods, algorithms are trained to make classifications or predictions, uncovering key insights within data mining projects. These insights subsequently drive decision-making within applications and businesses, ideally impacting key growth metrics. As big data continues to expand and grow, the market demand for data scientists will increase, requiring them to assist in the identification of the most relevant business questions and subsequently the data to answer them.

Deep Learning[2]: Deep learning attempts to mimic the human brain-albeit far from matching its ability-enabling systems to cluster data and make predictions with incredible accuracy.
Deep learning is a subset of machine learning, which is essentially a neural network with three or more layers. These neural networks attempt to simulate the behavior of the human brain—albeit far from matching its ability—allowing it to “learn” from large amounts of data. While a neural network with a single layer can still make approximate predictions, additional hidden layers can help to optimize and refine for accuracy. Deep learning drives much artificial intelligence (AI) applications and services that improve automation, performing analytical and physical tasks without human intervention. Deep learning technology lies behind everyday products and services (such as digital assistants, voice-enabled TV remotes, and credit card fraud detection) as well as emerging technologies (such as self-driving cars).

Tensorflow[3]: Tensorflow is an open-source software library for dataflow and differentiable programming across a range of tasks. It is a symbolic math library and is also used for machine learning applications such as neural networksKeras. It is used for both research and production by Google. Tensorflow is developed by the Google Brain team for internal Google use. It was released under the Apache License 2.0 on November 9, 2015. Tensorflow is Google Brain's
The second-generation system.1st Version of TensorFlow was released on February 11, 2017.
While the reference implementation runs on single devices, Tensorflow can run on multiple CPUs and GPU (with optional CUDA and SYCL extensions for general-purpose computing on graphics processing units). TensorFlow is available on various platforms such as 64-bit

Linux, macOS, Windows, and mobile computing platforms including Android and iOS. The architecture of TensorFlow allows the easy deployment of computation across a variety of platforms (CPU's, GPU's, TPU's), and from desktops - clusters of servers to mobile and edge devices. Tensorflow computations are expressed as stateful dataflow graphs. The name Tensorflow derives from operations that such neural networks perform on multidimensional data arrays, which are referred to as tensors.

American Sign language[4]: American Sign Language (ASL) is a complete, natural language that has the same linguistic properties as spoken languages, with grammar that differs from English. ASL is expressed by movements of the hands and face. It is the primary language of many North Americans who are deaf and hard of hearing and is used by many hearing people as well. ASL is a language completely separate and distinct from English. It contains all the fundamental features of the language, with its own rules for pronunciation, word formation, and word order. While every language has ways of signaling different functions, such as asking a question rather than making a statement, languages differ in how this is done. For example, English speakers may ask a question by raising the pitch of their voices and by adjusting word order; ASL users ask a question by raising their eyebrows, widening their eyes, and tilting their bodies forward.

Just as with other languages, specific ways of expressing ideas in ASL vary as much as ASL users themselves. In addition to individual differences in expression, ASL has regional accents and dialects; just as certain English words are spoken differently in different parts of the country, ASL has regional variations in the rhythm of signing, pronunciation, slang, and signs used. Other sociological factors, including age and gender, can affect ASL usage and contribute to its variety, just as with spoken languages. Fingerspelling is part of ASL and is used to spell out English words. In the fingerspelled alphabet, each letter corresponds to a distinct handshape. Fingerspelling is often used for proper names or to indicate the English word for something.

fig. 2.1

Long Short Term Memory[5]: An LSTM is a type of recurrent neural network that addresses the vanishing gradient problem in vanilla RNNs through additional cells, input and output gates. Intuitively, vanishing gradients are solved through additional *additive* components, and forget gate activations, that allow the gradients to flow through the network without vanishing as quickly.

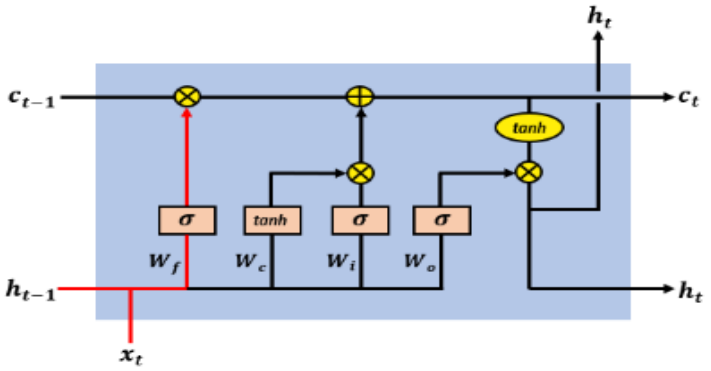

fig. 2.2

MediaPipe[6]: Real-time, simultaneous perception of human pose, face landmarks, and hand tracking on mobile devices can enable a variety of impactful applications, such as fitness and sports analysis, gesture control and sign language recognition, augmented reality effects, and more. MediaPipe, an open-source framework designed specifically for complex perception pipelines leveraging accelerated inference (e.g., GPU or CPU), already offers fast and accurate, yet separate, solutions for these tasks. Combining them all in real-time into a semantically consistent end-to-end solution is a uniquely difficult problem requiring simultaneous inference of multiple, dependent neural networks.

Tensorflow Model Garden[7]: The TensorFlow Model Garden is a repository with many different implementations of state-of-the-art (SOTA) models and modeling solutions for TensorFlow users. We aim to demonstrate the best practices for modeling so that TensorFlow users can take full advantage of TensorFlow for their research and product development.

LabelImg[8]: It is a graphical image annotation tool. It is written in python and uses Qt for its graphical user interface. Annotations are saved as XML files in PASCAL VOC format, the format used by ImageNet. Besides, it also supports YOLO and CreateML formats.

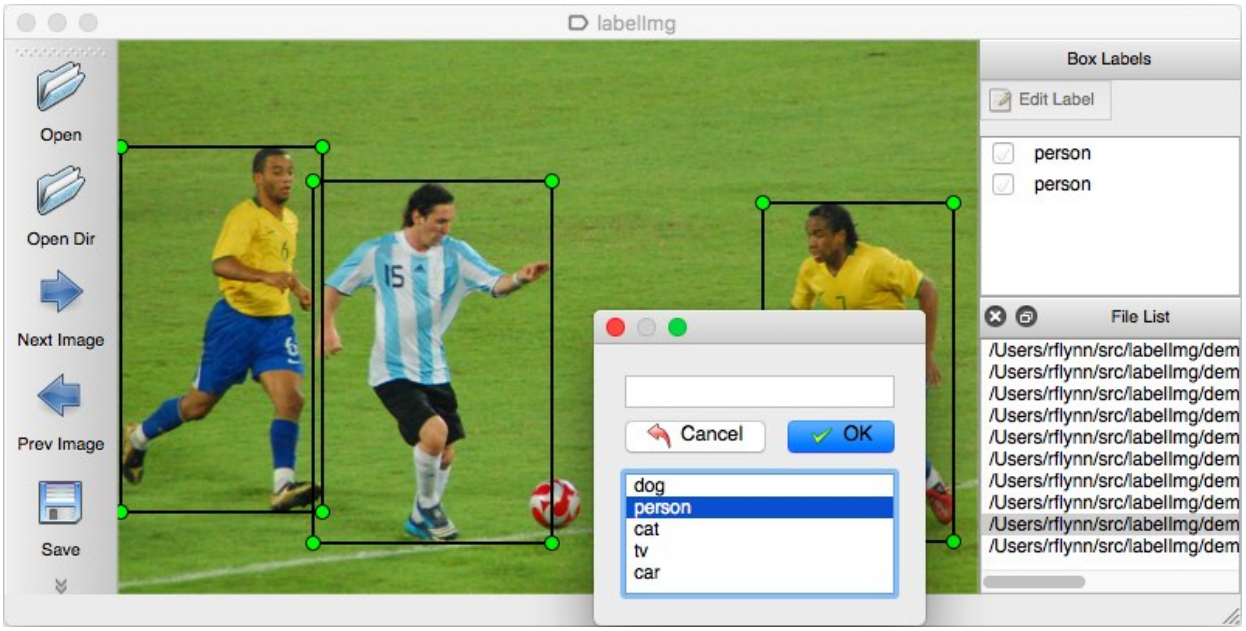

fig. 2.3

3. Requirement Specification

For the development of the project, the Iterative Model approach has been used as we are aware of the requirements and the project is open to changes and iterations in the future.

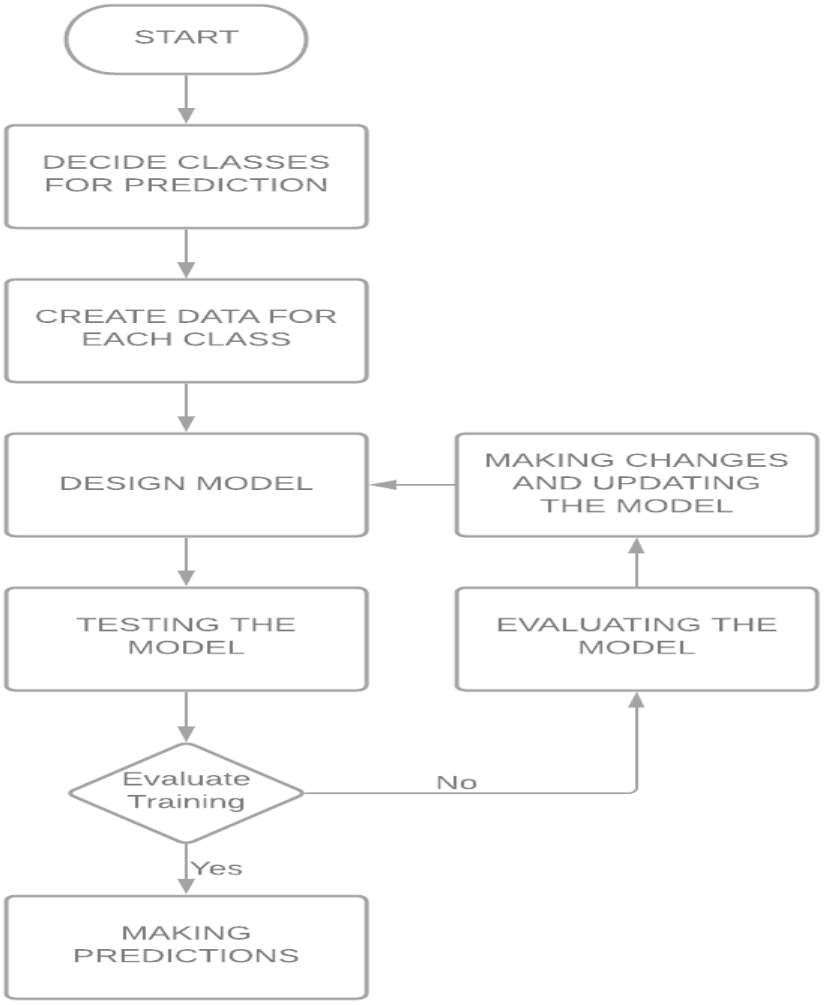

fig. 3.1

Software Development Life Cycle

Software Requirements:

- Jupyter Lab / Jupyter Notebook
- Python 3.7.4
- Iriun Webcam
- Tensorflow Board
- Tensorflow
- CUDA AND CUDNN (if GPU is available)

Hardware Requirements:

- RAM: 8 GB and above
- DISK SPACE: 8 GB and above
- Processor :
  - intel i3 10th Gen and above
  - intel i5,i7,i9 7th gen and above
  - AMD RYZEN 3000 series and above
- GPU (optional)

Constraints: For any latest generation computer/laptop having the above hardware specifications will have no problem running our project but the hardware of lower specification than mentioned above might provide some hindrance in the functioning of the model.

Supplementary Requirements: It is to be noted that these requirements are optional and their absence will not at all affect the performance of the model.

- GPU
- CUDA
- CUDNN
- Tensorflow GPU

The above mentioned should be installed according to the available GPU while making sure that the versions of CUDA, CUDNN, and Tensorflow GPU match with the pre-installed Tensorflow and the version of SSD Mobnet. Having these will provide a very smooth and fast-performing model.

4. Work Distribution

Responsibility	Starting Date	Ending Date	Member
Research about project	05/09/2021	20/09/2021	Vaibhav & Rashi
Dataset Creation	25/09/2021	15/10/2021	Vaibhav & Rashi
Labeling the dataset	25/10/2021	05/11/2021	Vaibhav & Rashi
Preparing the data for Training and testing	07/11/2021	10/11/2021	Rashi
Design Appropriate Model			Vaibhav
Testing & Evaluation			Vaibhav & Rashi
Design Webapp using Streamlit			Rashi
Prepare all necessary documents			Rashi

5. Design Document

Functional Description

1. **Title/Name: Hand Sign Detection**
2. **Purpose**

A. What is the application for and why is it necessary?

The Proposed application uses a TensorFlow model to detect various hand signs and translate their meaning to enable communication between mute and deaf people with the casual masses. The input will be taken in real-time using a webcam or a mobile camera and the frame will be passed to the trained model which will then detect and show the result.

B. Who uses it?

The proposed software can find its use at airports, railway stations, hotel receptions and in the future can also be deployed as a mobile app which will help in decreasing the gap between the communication of mute and deaf and normal people.

3. **Operating Environment:** The proposed software can be used as an offline or online application based on the circumstances with the availability of the specific resources.

4. **Functional Requirements:** The software takes a frame/image from the live feed coming from a webcam or mobile camera and passes it through SSD Mobilenet. SSD MobileNet V2 FPNLite 320x320 will compress the image to 320x320 in the pre-processing and in post-processing, it is going to take the detections that it found and convert it back to the original resolution. It uses Image augmentation, i.e. it might darken, shift or flip the image so that we can ideally get a better performing model.

5. **Performance Requirements**

- Tensorflow
- CUDA
- CUDNN
- GPU (Optional)

Above mentioned points are required for a smooth functioning model on the backend.

6. **Resource Dependencies**

- RAM: 8 GB and above
- DISK SPACE: 8 GB and above

- ➢ Processor :
  - intel i3 10th Gen and above
  - intel i5,i7,i9 7th gen and above
  - AMD RYZEN 3000 series and above
- ➢ GPU (optional)

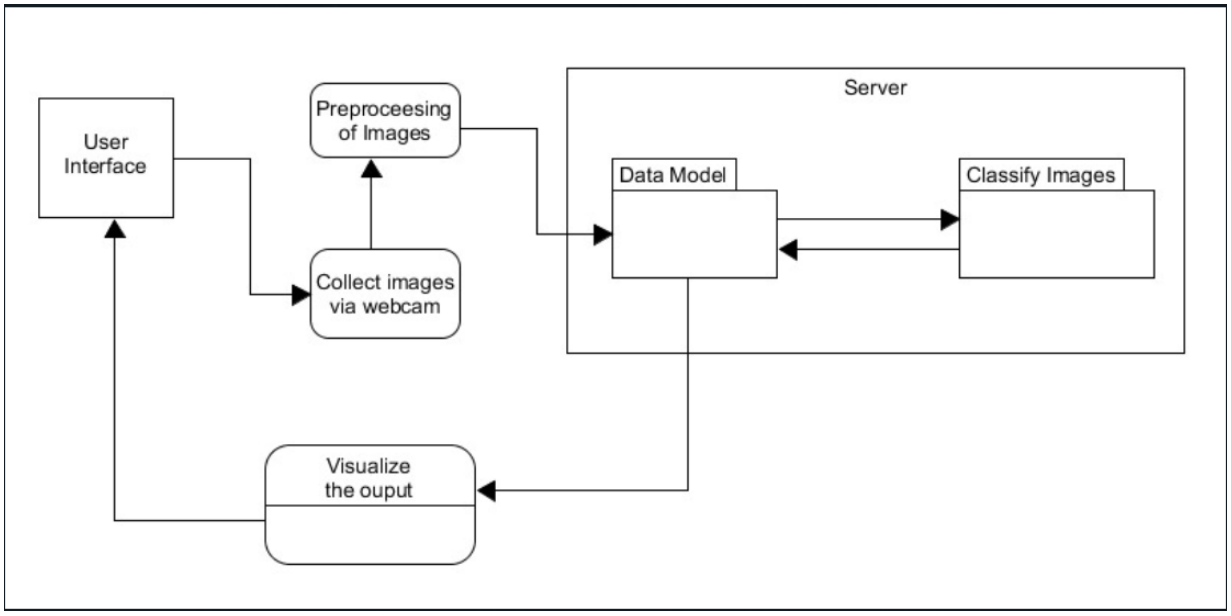

fig. 5.1 User Interface Design

Data Description

DATA: For training of the model images and their corresponding XML files containing the marked signs will be used. In layman terms there will be 2 sets of images and XML files; one will be used to train the model and one will be used for testing the accuracy of the model. The split ratio will be 80-20.

While performing real-time detections webcam/mobile camera will be used to provide the input where a frame will be taken from the live video and passed to the trained model for prediction While displaying the output accuracy of prediction will also be shown in percentage.

CPU: The application can be executed on any recent generation CPU and can function smoothly. The major processing will happen at the backend, the user only has to show signs to the webcam and the predictions will be visible instantly.

Access to Accelerator Hardware: Access to GPU will make the execution smooth with little to no issues for the future. SSD directly leverages CUDA and CUDNN with TensorFlow to perform huge calculations if need be and GPU speeds things up. The use of GPU will majorly be during the training of SSD however if need be it can also be leveraged during real-time predictions.

Access to Processed Data: The data is designed by the architects of the project from scratch so no third-party resource is used while doing so.

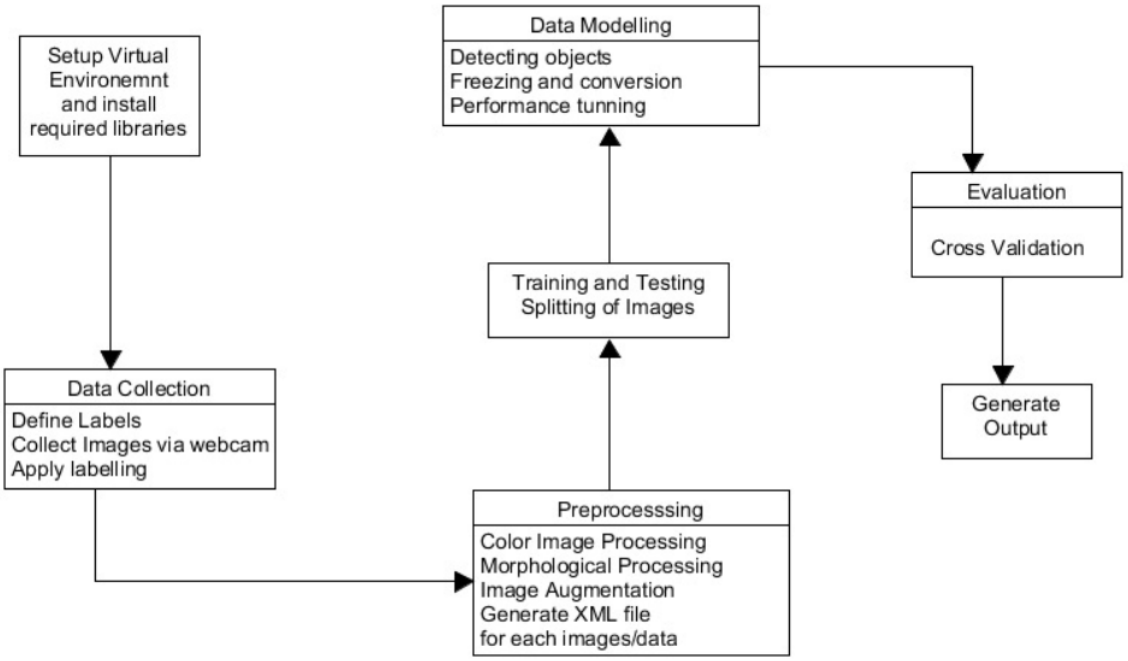

fig.5.2 Class Diagram

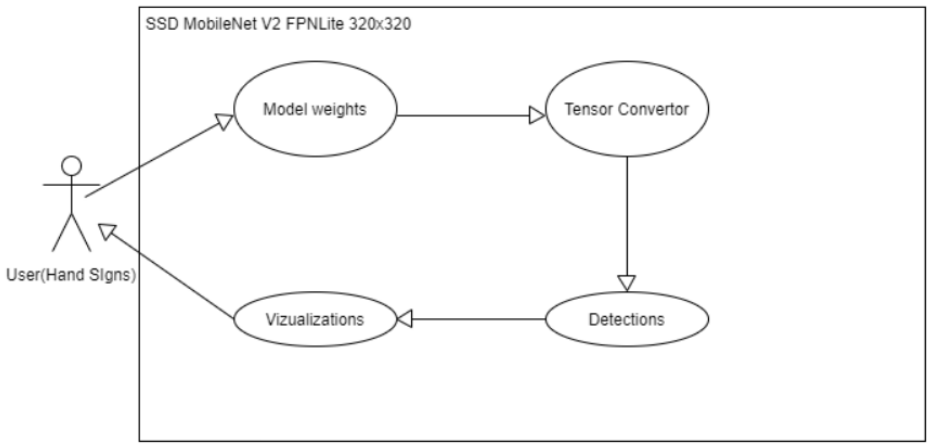

fig. 5.3 Use Case Diagram

6. Experimental Setup

Workflow of our project:-

1. Installing all the dependencies.
2. Defining labels to be collected and collecting data for each label.
3. Use LabelImg to label data and split data into test and train.
4. Train the model using SSD MobileNet V2 FPNLite 320x320.
5. Performing real-time detections and detection on images.
6. Freezing the graph and saving the model.

References

1. What is Machine Learning? - India
2. What is Deep Learning?
3. TensorFlow.org
4. What Is American Sign Language (ASL)?.
5. A Gentle Introduction to Long Short-Term Memory Networks by the Experts
6. MediaPipe Holistic — Simultaneous Face, Hand and Pose Prediction, on Device
7. tensorflow/models: Models and examples built with TensorFlow @misc{tensorflowmodelgarden2020, author = {Hongkun Yu and Chen Chen and Xianzhi Du and Yeqing Li and Abdullah Rashwan and Le Hou and Pengchong Jin and Fan Yang and Frederick Liu and Jaeyoun Kim and Jing Li}, title = {{TensorFlow Model Garden}}, howpublished = {\url{tensorflow/models: Models and examples built with TensorFlow}}, year = {2020}}
8. tzutalin/labelImg: 🖍 LabelImg is a graphical image annotation tool and label object bounding boxes in images