INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 07 ISSUE: 04 | APRIL - 2023 IMPACT FACTOR: 8.176 ISSN: 2582-3930

A Survey on Sign Language Recognition with Video Chat

Mrs. Kavyashree J
CSE dept.
Rajiv Gandhi Institute of Technology
Bangalore,India
kavya.jshankar@gmail.com

Arshiya M
CSE dept.
Rajiv Gandhi Institute of Technology
Bangalore,India
arshi9742@gmail.com

Mirza Akeel Abbas Baig
CSE dept.
Rajiv Gandhi Institute of Technology
Bangalore,India
akeelabbas29@gmail.com

Nida Haleema
CSE dept.
Rajiv Gandhi Institute of Technology
Bangalore,India
nidahaleema28@gmail.com

Noor Mohammed Masood
CSE dept.
Rajiv Gandhi Institute of Technology
Bangalore,India
mhdnoor06@gmail.com

Abstract—

The use of sign language is a crucial tool for improving communication between hearing-impaired people and the general public. SLR (Sign Language Recognition) systems in the past have been sophisticated and challenging to train. However, in this research, we provide a novel method that makes use of SSD MobileNet V2 FPNLite 320x320 pre-trained models and object recognition based on TensorFlow's object detection. By enabling the identification and detection of a set of images, this method streamlines the training process. The suggested system will be trained and evaluated using 10 to 15 different American Sign Language symbols. A fundamental social skill used to exchange information is communication. It is frequently used to express oneself and to fulfill fundamental human needs including the desire for protection, safety, and connection. Several stages, diverse methods, and distinct consequences are used in this procedure. It typically refers to a two-way exchange of information in the local vicinity of touch. Information flows far more easily when people are speaking the same language than when they are speaking languages that are from distinct language families. In order to facilitate video chat communication between signers and non-signers, our proposed sign language recognition system is specifically created. Each peer can see and hear the other during a video conversation thanks to their individual cameras and microphones. Nevertheless, using our method, a specific peer can also view the indicators that the other peer on the other end of the video chat exhibits or copies. Our system employs object detection to recognize and track the signer's hand motions in real-time in order to accomplish this. The technology then overlays a graphic of the detected sign onto the non-video signer's feed. This overlay is positioned so that it does not obstruct the view of the signer by the non-signer and is not obtrusive. By using contemporary technology, our suggested solution is made to enable distant and two-way communication between signers and non-signers. With the help of our system, video chat communication is improved, allowing signers and non-signers to communicate and interact more effectively.

Index Terms— SLR(Sign Language Recognition), Video Chat, Object Detection, SSD MobileNet.

1 INTRODUCTION

Broadly speaking, standardized signs or gestures enhance communication, as shown in army and aircraft security scenarios, interaction, human-machine systems, and telecontrol, to name a few. They also enhance efficiency, as seen in surgery. In particular, signals that are organized with syntax, semantics, grammar, pragmatics, morphology, and phonology create "sign languages," which enable us to convey thoughts and feelings in a manner similar to that of "natural languages." Globally, sign languages have spread. According to the World Health Organization, there are more than 466 million individuals who have hearing impairments, which is one reason why this happens (WHO, 2018). Muted, deaf, autistic, and hearing-impaired people, as well as their family members and teachers, use sign languages. Communication with those who are deaf or hard of hearing is unfortunately rather challenging because the majority of people are not accustomed to or are not familiar with sign languages. Fortunately, a growing number of technologies are devoted to resolving this complex problem by measuring indications and putting the proper meaning behind each action. There are many distinct sign languages used around the world for

communication, making them inconsistent. Examples include Indian Sign Language (ISL) in India and American Sign Language (ASL) in America. Our study on sign language understanding With Video Chat seeks to create a video chat assistive system that will instantly translate a signer's input into equivalent text during a video conference with a non-signer. The communication gap between the speech/hearing impaired and the rest of society can be closed with the aid of SLR devices. As a result, such systems provide a new avenue for applications based on human-computer interaction (HCI). The World Health Organization (WHO) research estimates that 278 million individuals worldwide were affected by hearing impairments in 2005. Its number increased by almost 14% to 360 million ten (10) years later. Since that time, the number has been rapidly rising. According to the most recent WHO report, 466 million individuals, or 5% of the world's population, had hearing loss in 2019. Of these, 432 million (or 83%) were adults and 34 million (or 17%) were children. By 2050, the WHO predicted that the population will double, reaching 900 million people. There is a need to remove the communication barrier that negatively impacts the lives and social interactions of these rapidly expanding deaf-mute populations. By converting sign language into spoken words and the other way around, sign language interpreters help close the communication gap with the hearing impaired. Although sign language has a flexible structure, there aren't enough qualified sign language interpreters available worldwide, which makes hiring interpreters difficult. Almost 70 million people use one of the more than 300 sign languages that exist, according to the International Federation of the Deaf. There is a need for a technology-based approach to replace traditional sign language interpreters in this situation. While communicating with someone using sign language, the upper body is used, including hand movements, facial expressions, lip reading, head nodding, and body postures. The most effective methods for understanding sign language are vision-based and wearable sensing modalities like sensory gloves. Systems for recognizing sign language based on these methods have been proposed by numerous researchers. In order to determine the hand posture for recognition, the glove-based system uses mechanical or optical sensors that are attached to the user's glove and transforms finger movements into electrical signals. In a vision-based technique, recognition is carried out utilizing object detection and features that correlate to the palms. In order to use this technique, the signs must be photographed or recorded on video, then processed using image-processing software.

2 LITERATURE SURVEY

This section aims to present an overview of the previous studies, methodologies, and techniques that are relevant to the current research problem. By critically analyzing the literature, this section aims to identify the research questions that need to be addressed and establish the rationale for the current research.

During our research, we have encountered five different literatures related to the topic of sign language recognition that we will be reviewing. The first paper, The researchers collected 100 samples of each American Sign Language (ASL) alphabet using gloves with sensors. They used an Artificial Neural Network to translate the signs into text with an accuracy of 99.8%. The gloves were low-cost and flexible with full-fiber. The model they used to detect the signs was called SSD Mobilenet V2, which extracts different features from the image. The process was fast and efficient. In another approach the researchers used a glove-based system to detect the bending of fingers and identify all the letters in the Polish Sign Language alphabet. A software uploaded in the microcontroller was used to identify the letters. The system had sign-to-text and sign-to-speech capabilities, and could also be used as a keyboard. They used computer vision to collect their dataset and detect the signs. Another group of researchers from the state of Bengal, India focused on 10 hand signs of the Bengali numerical sign language using a gesture-based, vision-based, sign-to-text approach. They used CNN and D-LBP for the detection process and preprocessed the data to detect the skin and signs in the picture. They used an ASL dataset and the model they used was the SSD Mobilenet V2 model. Another very peculiar paper we came across was one which was focused on gaming and VR experiences. In this paper the authors used 50 words and 20 sentences to develop an AI-based sign-to-text system using a triboelectric smart glove in a VR space. The system was designed to provide a VR-based game experience for deaf and mute individuals. They used computer vision instead of gloves and Kinect to develop a solution for day-to-day use. Another paper used the latest of the techniques that are new in the market which is Mediapipe library which is based on landmark detection. In this paper the researchers made a multilinguishtic recognition application worked on American alphabets and numbers, as well as Indian, Italian, and Turkish sign language, using a gesture-based, landmark-based, vision-based sign-to-text approach. They used the Support Vector Machine (SVM) algorithm along with the MediaPipe library. Their system had an average accuracy of 99% and was robust and cost-effective, requiring less computing power. They used landmarks and the MediaPipe library, which is a recent library with less research.

3 ARCHITECTURE

In this section, we present our Object detection-based neural network architecture for Real-Time SLR using SSD MobileNet V2 FPNLite 320x320. The flow diagram of the framework is depicted in Fig. 1, where the Camera Mounted system is used to acquire the sign inputs. Our application does not use gloves or sensors to capture hand movements. Instead, we utilize object detection techniques by manually marking and labeling the region of interest in images captured through a laptop or system camera using OpenCV library. The labeled images are then used to train a machine learning model that is capable of recognizing American Sign Language (ASL) symbols in real-time. We use the OpenCV library to capture images of 6 images per second. Our system allows for capturing images from a webcam or an inbuilt camera in a laptop. The user positions their hand within the frame and the system captures the image. The labeled images are then used to train a machine learning model. The extracted features were then stored and processed for prediction. Custom data is for example in our project we might take up to 10 or 15 different signs and we might take 30 images of each sign. Data pre-processing, we need to for example label each of our images as the symbol that they belong to. We will create a bounding box and mark its name. Load pre-trained model- we are using coco SSD mobile net320× 320 as a pre-trained model. This model has been trained on different images of the coco 2017 data set. This is capable of object detection on custom images. Training-training is nothing but where the machine learning actually happens the images are fed and the model is trained on those images. Features of the images are extracted and signs are able to be predicted. The export trained model- model has to be exported so that it can be used on a web page or a web application. Host trained model- the trained model has to be hosted in the cloud so that our web application with the use of the internet can load the model and make prediction real time. Load trained model- the trained model is loaded on the client side of the web application. The web application is based on socket.io for client server communicaton and peerJs for peer to peer communication. Our video chat application is loaded and the images from the video stream is captured at a rate of 1 image eper 16.7 milliseconds, that is because it suites the frame rates of most system displays. Data preprocessing - images from the camera module are to be modified to the form that is accepted by the machine learning model. Make a prediction- in this step, the predictions are to be rendered on the screen by creating a bounding box around the sign. And giving the confidence score and the saying which sign it is.

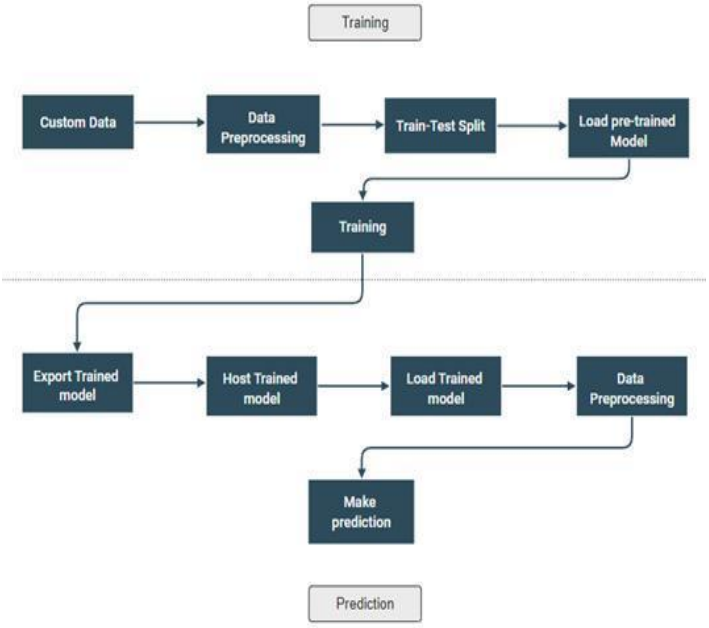
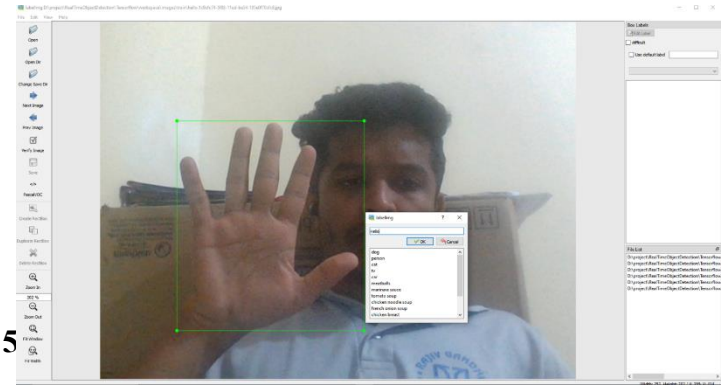


Fig 3.1 Block diagram of proposed System

SOME DAILY USE REPRESENTATION OF WORDS IN SIGN LANGAUGE

4 LABEL IMAGE

LabelImg is a graphical image annotation tool written in Python that uses Qt for its graphical user interface. Its primary purpose is to help users annotate images by drawing bounding boxes around objects of interest in the image.This dataset can be used to train and test machine learning models to detect signs in real-time video chat applications. LabelImg is designed to work with different image formats, such as JPEG, PNG, and BMP, and supports annotation in several formats, including Pascal VOC, YOLO, and TensorFlow Object Detection API. Users can easily navigate through multiple images and annotate them using keyboard shortcuts. The tool also allows users to save their annotations in a variety of file formats, including XML, JSON, and CSV.By using LabelImg, we were able to efficiently label our dataset, which is an essential step in training machine learning models for sign language detection LabelImg is an open-source tool, and its source code is available on GitHub for users to modify and improve. It is widely used by researchers, data scientists, and developers in the field of computer vision and machine learning to train and test their models.



5

Transfer learning is a machine learning technique where a model trained on one task is re-purposed for a different but related task. In the context of computer vision, it involves taking a pre-trained model that has been trained on a large dataset and using it as a starting point to train a new model on a smaller dataset.

In this case, "Pretrained model COCO SSD mobilenetV2 fpnlite320320" refers to a pre-trained object detection model. The COCO SSD (Common Objects in Context - Single Shot Detector) model is a deep learning architecture used for object detection tasks, while MobilenetV2 is a type of convolutional neural network (CNN) that is optimized for mobile and embedded devices. FPNlite is a lightweight feature pyramid network, which can extract useful features from images with different scales. The 320*320 refers to the input image size of the model.

One of the key benefits of transfer learning is that it allows you to leverage the rich knowledge and representations that a pre-trained model has learned from a large dataset. In the case of the COCO SSD mobilenetV2 fpnlite320*320 model, the model has been trained on the Common Objects in Context (COCO) dataset, which contains over 330,000 images and 2.5 million object instances across 80 object categories. By training on such a large dataset, the model has learned to detect a wide variety of objects in different contexts and under different conditions.

By using this pre-trained model as a starting point for a new object detection task, you can benefit from the features learned by the model during its training on a large dataset, without having to train the model from scratch on a small dataset. This can save a lot of time and computational resources.

You can fine-tune the pre-trained model by retraining the last few layers or adding additional layers to the model to adapt it to the new task. During training, the weights of the pre-trained layers are frozen, while the weights of the new layers are updated to minimize the loss function on the new dataset.

Overall, transfer learning with a pre-trained model like COCO SSD mobilenetV2 fpnlite320*320 can significantly improve the performance of an object detection model on a smaller dataset, as compared to training the model from scratch.

6 VIDEO CHAT

Web Real-Time Communication (WebRTC) is an open-source technology that allows real-time communication between browsers and mobile applications. Peer-to-peer communication is made available, allowing web browsers to exchange data, audio, and video. Video chat programmes are among the most used applications of WebRTC. It is the perfect option for real-time video chat apps since it provides a low latency and high-quality video and audio connection experience.

React JS is a popular JavaScript library used for building user interfaces. It provides a declarative syntax that makes it easy to build complex UI components. Integrating React with WebRTC enables developers to build real-time video chat applications with ease. With React, developers can create reusable UI components and manage the application state effectively. React also offers a virtual DOM, which makes it easy to update the UI in real-time, without

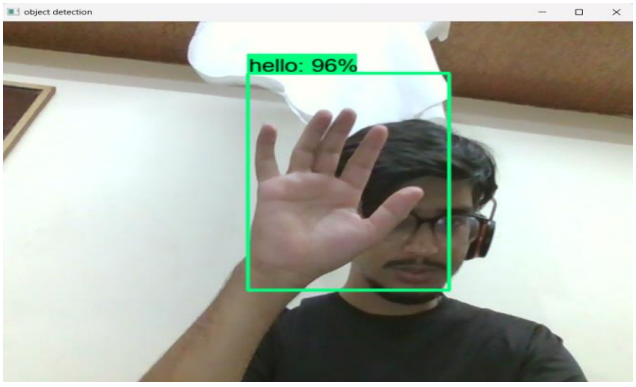
the need for reloading the page.

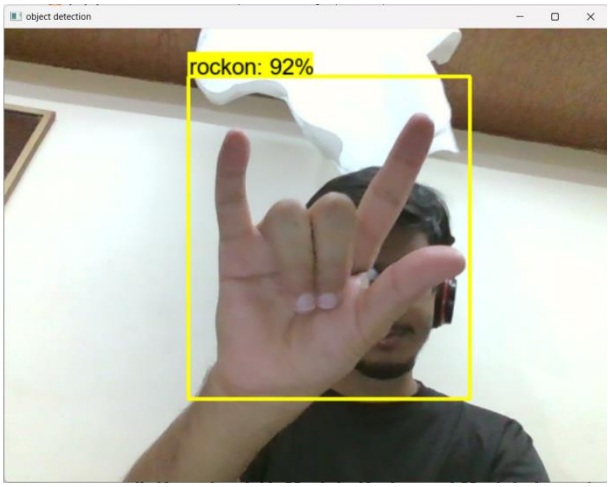
Sign language detection is an important feature in video chat applications that enable sign language interpretation. It allows people with hearing disabilities to communicate with others in real-time. Sign language detection uses computer vision algorithms to recognize sign language gestures and convert them into text or voice. Integrating sign language detection with WebRTC and React JS enables developers to build video chat applications that support sign language interpretation. With WebRTC, developers can capture video and audio streams, while React JS can be used to create the user interface and manage the application state. Sign language detection algorithms can be integrated into the application to recognize sign language gestures and provide real-time interpretation.

7 OUTCOME

Fully segmented words in a text are the system's anticipated output. Only a small word list will be used to train and classify this programme. The system can accurately read a gesture using a sensor that is a computer-mounted video camera.

The proposed method generates text-based output that aids in eliminating the communication gap between the deaf-dumb and average people. Systems for Sign Language Recognition (SLR) are designed to provide a quick and accurate method of translating sign language into text.





CONCLUSION AND FUTURE WORK

This study offers a comprehensive analysis of the various strategies and models. provides us with the high-level design and enables the software teams to make large-scale sketches and start prototyping. The terminology and other software lifecycle components are explained in modular documentation, which is beneficial. For large-scale machine learning and deep learning projects in particular, the tools and technologies used provide a potent library, tools, and resources for numerical computing. The order of the tasks helps us view the entire project plan in one location, work appropriately, and hopefully accomplish tasks on time.

REFERENCES:

[1]. Ronghui Wu, Sangjin Seo , Liyun Ma , Juyeol Bae , Taesung Kim “Full-Fiber Auxetic-Interlaced Yarn Sensor for Sign-Language Translation Glove Assisted by Artifcial Neural Network”(2022)ISSN 2311-6706.

[2]. Ewa Korzeniewska, Marta Kania and Rafał Zawi´slak “Textronic Glove Translating Polish Sign Language” Sensors 2022, 22, 6788.

[3]. F. M. Javed Mehedi Shamrat , Sovon Chakraborty, Md. Masum Billah , Moumita Kabir , Nazmus Shakib Shadin , Silvia Sanjana “Bangla numerical sign language recognition using convolutional neural networks” Indonesian Journal of Electrical Engineering and Computer Science Vol. 23, No. 1, July 2021, pp. 405~413 ISSN: 2502-4752.

[4]. Songyao Jiang§ , Bin Sun§ , Lichen Wang, Yue Bai, Kunpeng Li and Yun Fu Northeastern University, Boston MA, USA "Skeleton Aware Multi-modal Sign Language Recognition" 2021 W911NF-17-1-0367.

[5]. Feng Wen, Zixuan Zhang, Tianyiyi He & Chengkuo Lee "AI enabled sign language recognition and VR space bidirectional communication using triboelectric smart glove" (2021) 12:5378.

[6]. Ozge Mercanoglu Sincan, Julio C. S. Jacques Junior, Sergio Escalera, Hacer Yalim Keles "ChaLearn LAP Large Scale Signer Independent Isolated Sign Language Recognition Challenge: Design, Results and Future Research"(2021).

[7]. Ilias Papastratis , Kosmas Dimitropoulos and Petros Daras "Continuous Sign Language Recognition through a Context-Aware Generative Adversarial Network" Sensors 2021, 21, 2437.

[8]. Arpita Haldera , Akshit Tayadeb "Real-time Vernacular Sign Language Recognition using MediaPipe and Machine Learning" International Journal of Research Publication and Reviews Vol (2) Issue (5) (2021).

[9]. Lu Meng and Ronghui Li "An Attention-Enhanced Multi-Scale and Dual Sign Language Recognition Network Based on a Graph Convolution Network" Sensors 2021, 21, 1120.

[10]. Giovanni Saggio, Pietro Cavallo , Mariachiara Ricci ,Vito Errico , Jonathan Zea and Marco E. Benalcázar "Sign Language Recognition Using Wearable Electronics: Implementing k-Nearest Neighbors with Dynamic Time Warping and Convolutional Neural Network Algorithms"Sensors 2020, 20, 3879.