IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 5, May 2024

SIGNSense: Auditory -Optic Impairment Communication Bridge

B H Theja[1], Harshitha S[2], Likitha G[3], Dr. Soumya Patil[4]

B.E. Students, Department of Computer Science & Engineering[1,2,3]

Associate Professor, Department of Computer Science & Engineering[4]

Sir M Visvesvaraya Institute of Technology, Bengaluru, India

Abstract: *Language experts have recognized sign languages as natural languages with the ability to convey human emotions and ideas. Translation from written language into sign videos or extraction of spoken language sentences from sign videos is the aim of sign language translation. Sign language is the principal means of communication for the deaf and hard of hearing community, which comprises 32 million children and 328 million adults worldwide who suffer from hearing impairment. However, the inability of current systems to accurately translate and transmit sign language motions in real-time prevents effective and spontaneous communication. This research provides a revolutionary technique that enables real-time recognition of ISL gestures by integrating natural language processing with cross-modal integration. The methodology uses cutting-edge methods like the Single Shot Multibox Detector (SSD) with MobileNetV2 architecture for data collection, preprocessing, model selection, and training. In real-time inference, the trained model attains an impressive 94% accuracy rate, showcasing strong performance and encouraging outcomes for enhancing communication accessibility for people with hearing impairments*

Keywords: Sign Language, ISL gestures, SSD, MobileNetV2

I. INTRODUCTION

Sign languages have been proven linguistically to be natural languages, just as capable of expressing human thoughts and feelings as traditional languages are.[7] The goal of sign language translation is to either convert written language into a video of sign (production) or to extract an equivalent spoken language sentence from a video of someone performing continuous sign.[8] According to the World Health Organization, there are 328 million people worldwide suffer from impaired hearing loss, of whom 32 million are children. Sign language is the daily language of communication between deaf and dumb people, which is the most comfortable and natural way of communication between deaf and dumb people, and is also the main tool for special education schools to teach and convey ideas. [9] The signs of Indian Sign Language (ISL) are as illustrated in Figure 1. ISL is a complete language with its own syntax, grammar, lexicon, and several other unique linguistic characteristics. Over five million people who are deaf use it in India. Hand gesture recognition is the categorization of hand motions into discrete groups according to their style. The technique of categorizing hand motions based on their style is known as hand gesture recognition. Signal processing, artificial intelligence, and statistics are only a few of the many methods that we utilize for computation. For the process of recognizing images obtained from the Kinect sensor, one of the algorithms utilized is 3D-CNN. 3D-CNN convolves 3D data using 3D filters, then sends the resulting convoluted data to pooling layers before to the fully connected layers for classification.

The inability of current technologies to quickly and accurately translate and transmit sign language motions makes it more difficult for people with dual sensory impairments to communicate spontaneously and effectively.Real-time communication is challenging with existing systems since they frequently have latency problems. For the system to be practically useful, all hardware (such as cameras and sensors) and software (such as online platforms and mobile applications) must be seamlessly integrated. Through tackling these issues, this article seeks to develop a game-changing solution that improves the communication skills of people with visual and aural impairments, promoting more accessibility and inclusivity in their day-to-day lives.

DOI: 10.48175/IJARSCT-18466

Our primary goals are natural language processing and cross-modal integration. We aim to investigate the intersection of natural language processing and cross-modal integration in the context of sign language recognition in this work. By developing a system that can accurately identify Indian Sign Language gestures in real time, the project hopes to improve communication for those who struggle with hearing loss. Gathering a comprehensive and representative set of Indian Sign Language gestures is the first stage of the research. Before being utilized for training, this dataset is first preprocessed to enhance image quality, including shrinking and normalization

Using OpenCV, we capture video frames from a camera during the real-time inference phase. Every frame undergoes preprocessing to make sure it matches the input format required by the trained model. After the trained model has been used to predict the expected sign language gesture for each frame, it is then displayed in real time on the video feed. The process of analysis includes evaluating the system's accuracy and speed. Furthermore, we identify and resolve any challenges or limitations observed during the real-time inference, such as hand orientation.or lighting. By developing a system that can accurately identify Indian Sign Language movements in real-time, the main purpose of this research is to improve the field of sign language recognition and provide better communication for deaf individuals.

II. LITERATURE SURVEY

Using MATLAB-implemented Artificial Neural Networks (ANN) and Principal Component Analysis (PCA), Kusumika Krori Dutta and Sunny Arokia Swamy Bellary employed Sunny machine learning approaches.The categorization of Indian Sign Language (ISL) movements is thoroughly examined in the [1] paper, with an emphasis on both single- and double-handed gestures. It draws attention to how important communication is for people with disabilities who predominantly express themselves through sign language.The training protocol is described in the study, along with the use of machine learning techniques that facilitate supervised learning for gesture identification, such as Back Propagation and K Nearest Neighbors (K-NN).

To assist individuals "hear" sign language, the authors of [2] article have developed and implemented DeepSLR, a real-time end-to-end continuous SLR system that translates sign language into voices. To accurately record the arm movements and fine-grained finger motions, both sMEG and IMU sensors are used. A multi-channel CNN and an attention-based encoder-decoder architecture were presented to achieve accurate, scalable, and continuous SLR from beginning to end without the need for sign segmentation.

The only extensive publicly accessible dataset on Indian Sign Language (ISL) is INCLUDE, which is presented in the [3] publication. Deep models for Sign Language Recognition on ISL can be explored thanks to the dataset's quantity and quality. In their evaluation of several deep learning models, they find one that achieves a high accuracy. They also show that our approach on an American Sign Language dataset achieves state-of-the-art accuracy. This model solely trains the decoder; the feature extractor and encoder are pre-trained. By optimizing the decoder for an American Sign Language dataset, they investigate generalization even more.

In [4], Nantinee Soodtoetong and Eakbodin Gedkhaw describe a deep learning method for gesture detection in sign language. They use 3D-CNN to detect motion and a Kinect sensor to capture RGB-D images. They then assess the motion recognition system's effectiveness using five different words. The outcomes demonstrate that the 3D-CNN system was able to accurately identify the gesture motion.

The [5] paper describes a unique method for identifying gestures in Indian Sign Language (ISL) using convolutional neural networks (CNN) on continuous sign language videos taken in selfie mode with a mobile application. Stochastic pooling is found to be the best appropriate technique for this application after several CNN architectures are investigated and tried on the dataset.The rectified linear units (ReLu), stochastic pooling layers, a softmax output layer, and four convolutional layers make up the suggested CNN design. This has shown better results than classical classifiers like the Mahalanobis distance classifier (MDC).

Leap Motion Controller (LMC) is a 3D motion sensor that can be used to construct a system for American Sign Language (ASL) recognition, as discussed in the paper [6]. A Multi-Layer Perceptron (MLP) neural network with Back Propagation (BP) algorithm is used for classification, whereas LMC is used for data collecting and hand gestures are used for feature extraction in the proposed system.

III. METHODOLOGY

Data collection:

OpenCV in Python and a webcam were used to capture images. Seven hundred pictures in all, one hundred pictures for each class (A, B, C, D, E, G, and T). Real-time image capturing was done, and each class had a folder for storing photos. The suggested approach for recognizing sign language is predicated on a webcam frame that was taken with a laptop or PC.Image processing is done with the OpenCV Python computer library. To improve accuracy through a huge dataset, several photographs of different sign language symbols were collected under varied lighting circumstances and from different perspectives.

Data Preprocessing:

To enhance model performance, the photos were preprocessed by shrinking them to a consistent size (such as 320x320 pixels) and normalizing pixel values.Rotation, flipping, and brightness adjustment were added to the dataset to boost the training data's variability.

Model Selection and Configuration:

Because of its ability to balance speed and accuracy, the SSD (Single Shot Multibox Detector) model with MobileNetV2 as the basis architecture was chosen. The TensorFlow Object Detection API has produced a pre-trained object detection model, ssd_mobilenet_v2_fpnlite_320x320_coco17_tpu-8. With MobileNetV2 serving as its foundation network, it is built on the Single Shot Multibox Detector (SSD) architecture. The COCO (Common Objects in Context) dataset is used to train the model, and Tensor Processing Units (TPUs) are the target platform for optimization.This model is appropriate for applications where efficiency is crucial, like embedded systems or mobile devices, because it is made for real-time object detection jobs requiring comparatively little CPU power. The label "320x320" denotes the expected input image resolution of 320 pixels in width and 320 pixels in height, as predicted by the model. The photos are labeled using this software. two output classes (background and hand gesture) were added to the model's configuration, and its hyperparameters (learning rate, batch size, and number of training steps) were changed in response to performance testing and experimentation.

Image Labelling:

An excellent resource for annotation of the Indian Sign Language (ISL) dataset used in this study was LabelImg, a widely used graphical image annotation tool. The tool made it easy for people to manually identify the item bounding boxes around ISL hand motions, ensuring that the dataset was properly annotated for training the sign language recognition model. The annotated dataset provided crucial ground truth labels during the training phase, which enabled the model to recognize the distinct characteristics of each ISL motion. Because of its intuitive interface and support for numerous annotation formats, LabelImg proved to be a valuable tool for this study, expediting the annotation process and ensuring the dataset's accuracy. Its compatibility with multiple operating systems raised its usefulness even more, making it an essential component in the creation of the real-time ISL gesture recognition system.

Model Training:

Using weights pre-trained on a sizable dataset (such as the COCO dataset), the model was initialized to take advantage of the characteristics that had been learned. The pre-trained model was adjusted using the ISL dataset to better suit the demands of the particular hand gesture detection job. Specify the model directory, pipeline configuration file, and number of training steps when configuring the training environment using the TensorFlow Object Detection API.During the training process, the model's performance was evaluated by keeping an eye on the validation metrics and loss function. The model was trained with a batch size of 16 and set to detect 10 classes, including background. With a batch size of 10, the training was conducted over a total of 20000 steps. To enhance the model's performance on the particular task of hand symbol detection, the weights were optimized throughout the training phase.

DOI: 10.48175/IJARSCT-18466

Model Evaluation:

To verify the model's efficacy in real-time sign language recognition, a test dataset distinct from the training data was used. Accuracy, precision, recall, and F1-score—all often employed in classification tasks—were among the evaluation measures. The model demonstrated its ability to accurately classify ISL motions with an outstanding 94% accuracy rate. There were few false positive detections, as indicated by the high precision, which is the ratio of true positive predictions to all positive predictions. Recall was likewise high, suggesting that the model was able to identify a sizable fraction of real ISL gestures. Recall is defined as the percentage of genuine positive predictions among all actual positive cases. To provide a balanced assessment of the precision and recall, the F1-score—the harmonic mean of both—was also computed. Overall, the evaluation findings show how well the model works in real-time at reliably identifying ISL gestures, underscoring its potential for useful applications in enhancing the accessibility of communication for those with hearing impairments.

Detection:

To initialize the model for inference, it imports the pre-trained detection model that was given in the pipeline config file and restores the checkpoint. Using an image as input, this function preprocesses it, uses the detection model to do inference, and then post-processes the output to obtain the final detections.

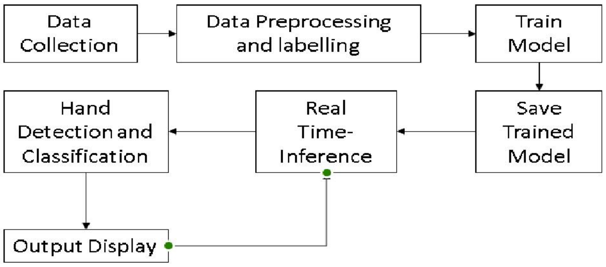

Figure 1 Sign Language Detection Approach

IV. SOFTWARE AND TOOLS

The project was implemented in Python using TensorFlow and OpenCV. The version of TensorFlow used was 2.10.1, and OpenCV4.4.0 was used for image processing and data collection.

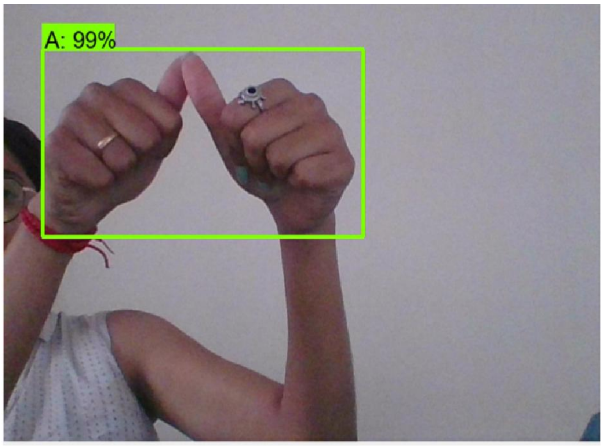

Fig 1. Detection of alphabet A

DOI: 10.48175/IJARSCT-18466

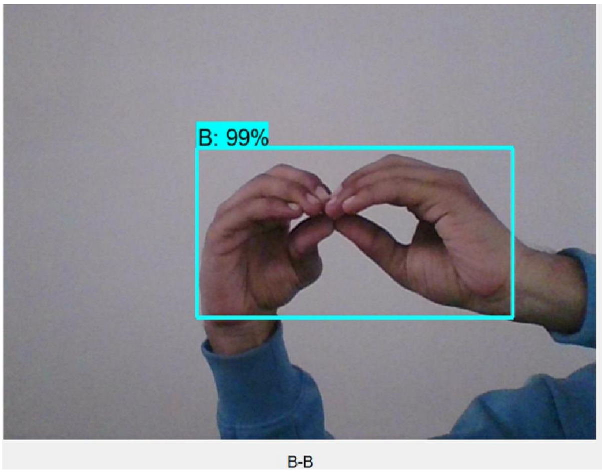

Fig 2. Detection of alphabet B

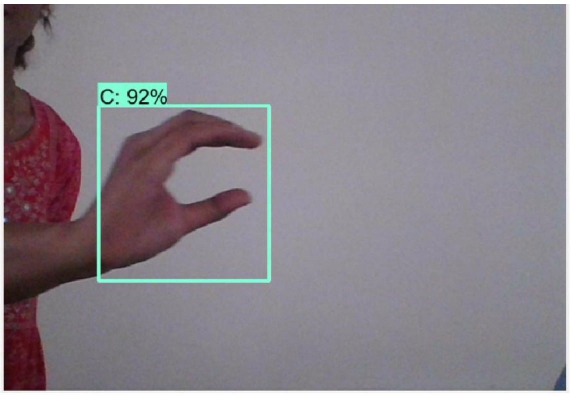

Fig 3. Detection of alphabet C

V. RESULT

The research project aimed to develop a real-time sign language recognition system, specifically focusing on the Indian Sign Language (ISL). The project successfully collected and pre-processed a dataset of ISL gestures, consisting of 100 images for each of the 7 classes, and trained a model using transfer learning with a pre-trained MobileNetV2 base. The trained model demonstrated robust performance in real-time inference, accurately recognizing ISL gestures from webcam video frames. The system's performance was evaluated in terms of speed and accuracy, showing promising results for improving communication accessibility for individuals with hearing impairments. Challenges such as lighting conditions and hand orientation were addressed, enhancing the system's robustness. The accuracy of the model was measured at 94%, highlighting the effectiveness of the proposed approach in real-time ISL gesture recognition and showcasing its potential for practical applications in inclusive communication technologies.

REFERENCES

[1]. Kusumika Krori Dutta and Sunny Arokia Swamy Bellary, “MachineLearning Techniques for Indian Sign Language Recognition”,Sept2017,doi:https://doi.org/10.1109/CTCEEC.2017.8454988

[2]. Zhibo Wang, Tengda Zhao, Jinxin Ma, Hongkai Chen, Kaixin Liu, Huajie Shao, Qian Wang, Ju Ren, “Hear Sign Language: A Real-time End-to-End Sign Language Recognition System” Dec 2020, doi:10.1109/TMC.2020.3038303.

[3]. Advaith Sridhar, Rohith Gandhi Ganesan, Pratyush Kumar and Mitesh Khapra, “INCLUDE: ALarge Scale Dataset for Indian Sign Language Recognition,” Multimedia (MM’20), October 2020, Seattle, WA, USA. ACM, New York, NY, USA, https://doi.org/10.1145/3394171.3413528

[4]. Nantinee Soodtoetong and Eakbodin Gedkhaw, "The Efficiency of Sign Language Recognition using 3D Convolutional Neural Networks",July2018,doi:https://doi.org/10.1109/ECTICon.2018.8619984

[5]. G.Anantha Rao, K.Syamala, P.V.V.Kishore and A.S.C.S.Sastry, " Deep Convolutional Neural Networks for Sign Language Recognition",Jan2018,doi:https://doi.org/10.1109/SPACES.2018.8316344

[6]. Deepali Naglot and Milind Kulkarni, "Real Time Sign Language Recognition using the Leap Motion Controller", August 2016,doi: https://doi.org/10.1109/INVENTIVE.2016.7830097

[7]. Beifang Yi, Xusheng Wang, Frederick C. Harris, Jr, and Sergiu M. Dascalu, "sEditor: A Prototype for a Sign Language Interfacing System" Transactions on Human-Machine Systems.,Vol.44,No.4,August2014,doi:https://doi.org/10.1109/TSMC.2014.2316743

[8]. Necati Cihan Camg¨ oz , Oscar Koller , Simon Hadfield and Richard Bowden, "Sign Language Transformers: Joint End-to-end Sign Language Recognition and Translation", Computer Vision and Pattern Recognition June2020, doi:https://doi.org/10.1109/CVPR42600.2020.01004

[9]. Siming He, "Research of a Sign Language Translation System Based on Deep Learning", Artificial Intelligence and Advanced Manufacturing,October2019,doi:https://doi.org/10.1109/AIAM48774.2019.00083

DOI: 10.48175/IJARSCT-18466