Real-time Sign Language Detection

Prof. Archana Chaudhari [1], Harsh N. Chandak[2]

[1,2] Department of Instrumentation and Control

BRACT's Vishwakarma Institute of Technology

Pune, India

Email: archana.chaudhari@vit.edu [1], harshnchandak@gmail.com[2]

***Abstract* —** To facilitate communication even when a person does not understand sign language, a system or application that can recognize sign language gestures is required. With this work, we hope to make a step in using sign language recognition to close this communication gap. In this report, we used and tested three different approaches for developing a system to recognise sign language, the first two of which were deep learning approaches, one of which involved the use of a convolutional neural network, and the third a machine learning approach that involved the use of a feed-forward neural network. And the training accuracy was above 98% for all three but the machine learning method has higher accuracy in real-time than rest two.

***Keywords*—**Sign language, Convolution neural network (CNN), feed-forward neural network, deep learning, machine learning.

1. INTRODUCTION

The deaf have traditionally used sign language to communicate. However, because sign language is not frequently employed, there is still a communication gap. In order to distinguish different ASL signals in real-time, we sought to employ machine learning and deep learning in this project. There are several signals for the words and expressions we use every day in American sign language (ASL), which is commonly used. Additionally, if a model exists that can instantly understand these signs, communication barriers will be severely reduced and easier to bridge. To solve this challenge, we applied deep learning and machine learning techniques in this work. We were able to recognize the signs in real-time using Python by utilizing TensorFlow and OpenCV. We initially used CNN to train the model and tried the deep learning approach. However, there was still a need for improvement, so after continuous testing and study, we decided to apply a feed-forward neural network. Likewise, the accuracy was achieved with a high-speed detection rate using the feedforward neural network and the hand detection API of the mediapipe python library.

2. EXISTING LITERATURE

For this issue statement, there are several projects available, although the majority of them concentrate on recognizing the alphabets and numerals used in sign language, or some programs identify their local dialect.

The first paper we read was Siming He's paper [1]. He suggested a program with a database of 10,000 sign language graphics and 40 often used terms. Using Faster R-CNN with an incorporated RPN module, locate hand regions in the video frame. Performance is improved by increased accuracy. Compared to single-stage target acquisition strategies like YOLO, segmentation can be completed more quickly. On paper, when compared to Fast-RCNN, Faster R-CNN acquisition accuracy rises from 89.0 to 91.7 percent. Rekha, J.'s research [2] also used the YCbCr skin model to identify and distinguish the skin of the touch skin. They applied the bending concept and extracted picture elements and were separated by Multi-class SVM, DTW, and indirect KNN. 25 images were used for testing and 23 static letter signs in Indian Sign Language were used for the training dataset. 94.4 percent for static and 86.4 percent for dynamic were the experimental results.

Real-Time Recognition of Sign Language Gestures (Words) from Video Sequences [3]. By Sarfaraz Masood, Adhyan Srivastava, Harish Chandra Thuwal, and Musheer Ahmad using CNN and RNN. The authors employed the deep convolutional neural network (CNN) inception model to train the model on spatial features, and they used recurrent neural networks (RNNs) to train the model on temporal features. Argentinean Sign Language (LSA) movements from 46 different categories make up our dataset. The suggested model was

successful in reaching a high accuracy of 95,2 percent across a large collection of pictures. Let's now examine the work of Rung-Huei Liang and Ming Ouhyoung. [4]. They created a prototype system with a lexicon of 250 words in Taiwanese Sign Language (TWL). For the 51 basic postures, 6 orientations, and 8 actions. In this system, Hidden Markov Models (HMMs) are used. 80.4 percent of them were detected on average. The next one is finished by Mehreen Hurroo and Mohammad Elham[5]. They developed a system to recognise sign language, and it was having 90% accuracy at identifying 10 American sign gestures by combining computer vision and convolutional neural networks.

This paper[6] uses techniques like feature extraction and thresholding, and with the use of these features, they proposed a simple method for number recognition. For number recognition, they used thresholding values. They break down that process into three stages: first, they used a web camera to record an image; second, they applied a threshold value to the image; and third, they used the threshold value to identify the digits. This work[7] presented a dynamic hand motion detection system for home appliance management using only the depth camera. Static hand postures and hand trajectories are used to identify the dynamic hand gesture. Seven frequently used dynamic hand motions can be recognized by the suggested approach. The technology is effective for controlling domestic appliances, according to experimental data. This paper[8] discusses a fundamental recognition method that uses thresholding to identify integers between 0 and 10. The overall algorithm consists of three major steps: image capture, threshold application, and number recognition. The concept is put forth that the user must use coloured hand gloves.

3. METHODOLOGY

As mentioned earlier this project was done in three different methods.

3.1. Method 1 –

So, this method was more of predicting the comparison between the dataset and the webcam feed rather than the prediction of the signs. This approach can be separated into three sections.

Data collection – Data collection is an important part of any system as the performance of the system is dependent on it. For this model, the images were captured using a webcam a total of 75 images were captured out of which 80% were for training and 20% were for testing. The images were collected using the same background and were written in jpg format.

Data labelling – After the images were captured, they were labelled for each sign using the label-image package in python. The labelling data is then stored in .xml format which is later used for model training.

Model Training and detection – Using neural network model training was done with the help of TensorFlow API and Keras module of python. After the training was done the checkpoint for the training is generated using detection in the real-time was done with the help of OpenCV python.

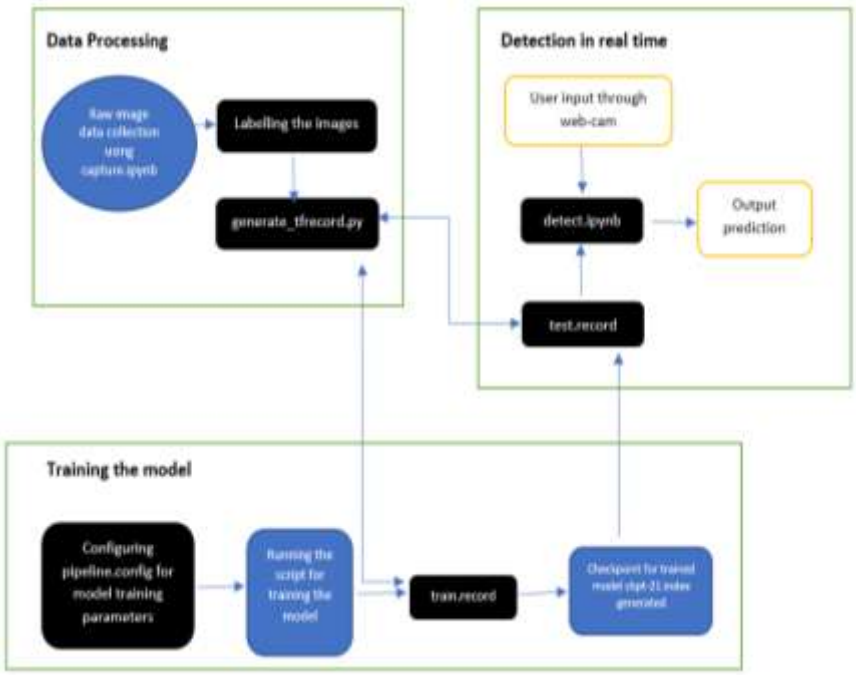

Fig.1. System Architecture for method A.

Limitations – There were many limitations to this method. The model was able to detect the signs only in front of the background which was used to capture image to create dataset. Also, it wasn't really detecting it was simply comparing the web cam feed with the dataset. Also, this system was very slow.

3.2. Method 2 –

This was the advanced model training involving CNN. In this approach we used a 5 layered CNN. This method has also 2 parts.

Data collection and augmentation – Data collection is done using capturing the hand signs using web cam. After that the images were separated to their individual classes. Then the by performing augmentation on the dataset size of the dataset was increased and further this augmented dataset was used for model training.

Model Training and CNN – After the augmentation a 5 layered CNN was applied. First layer was input layer. Second layer is conv2D layer, then there are 2 hidden layers for parsing and pooling and last one is output layer. For the training MobileNet was the base model used from keras API. Training accuracy achieved was 98%.

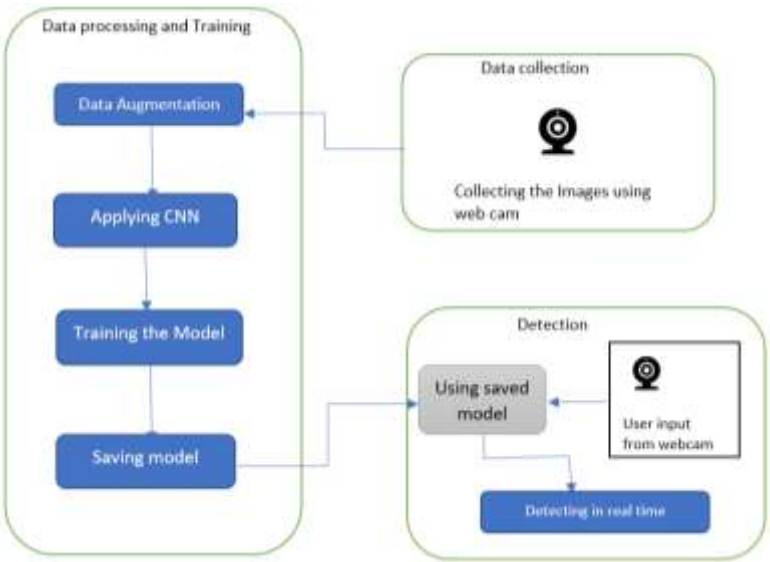

Fig.2. System Architecture for method B

Limitations – Although this system was fast there was one drawback of this system it was also predicting the other signs. And predictions are also highly dependent on the lightening and the background.

3.3. Method C –

This method was a machine learning approach. In this method pre-processing was easily done using Mediapipe library of python as it has an inbuilt API for detecting the hands. Steps for this method were

Importing Mediapipe and Collecting Data – Mediapipe module detect the hands by plotting 30 landmarks or the key points on the hand.

Fig.3. hand key point marking using mediapipe

Then before collecting the data Normalization was done. Mediapipe stores hand data in a form of 2-D array and for training our model we need to convert that 2-D array in a one-dimensional array. And for that purpose normalization was done. After the normalization was done the data was stored in the .csv file marked with a key assigned for that specific sign ready for feeding it to our neural network for training.

Model training and the neural network – After the dataset creation, next step was the model training. For the model training feed forward neural network approach was used. In this method training accuracy was 99.5% and while detecting in the real time model was able to detect all the signs instantaneously and accurately.

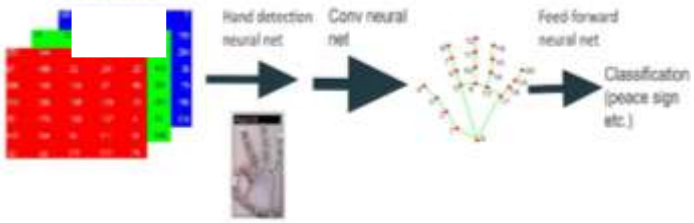

Fig.4. feed forward neural network.

We applied a 4 layered neural network to our system, in which first and last layers are input and output layers respectively and there are two hidden layers for parsing and pooling.

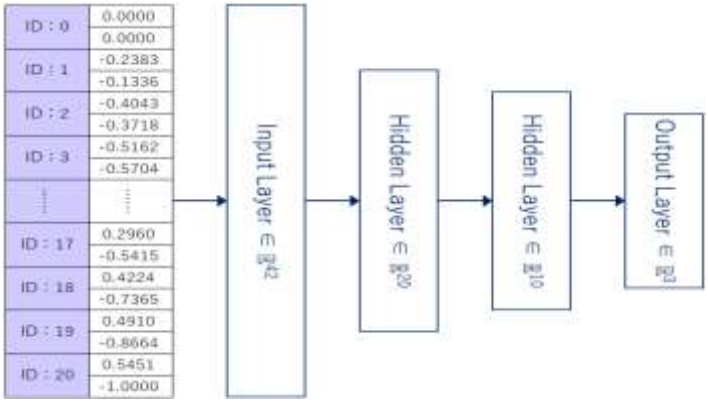

Fig.5. structure of the model prepared.

This model has the higher accuracy compared to the other two, below is the confusion matrix :

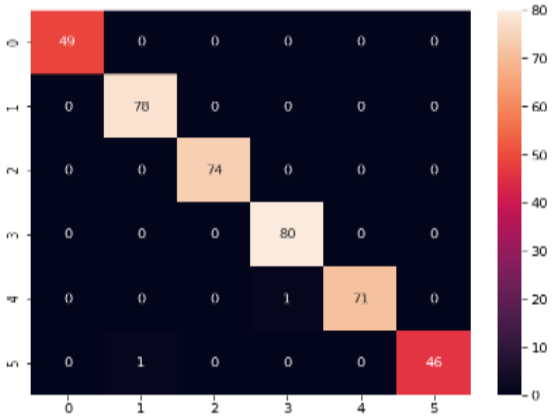

Fig.6. confusion matrix

Table 1. Classification Report

Sr. No	Precision	recall	F1-score	support
0	1.00	1.00	1.00	49
1	0.99	1.00	0.99	78
2	1.00	1.00	1.00	74
3	0.99	1.00	0.99	80
4	1.00	0.99	0.99	72
5	1.00	0.98	0.99	47
Accuracy			0.99	400
Macro avg.	1.00	0.99	0.99	400
Weighted avg.	1.00	0.99	0.99	400

As we can see accuracy was > 99%.

4. RESULTS AND DISCUSSION

Table 2. comparison of all three Methods

Methods	**Speed**	**Training Accuracy**	**Detection Precision**	**Effect of the surrounding**
Method A	Slow	97%	Very low	Yes
Method B	Fast	98%	Low	Yes
Method C	Fast	99.5%	Accurate	No

This is the comparison table for all the three methods mentioned earlier in this paper. The results for the machine learning approach i.e. method 3 are higher and more accurate than remaining two and detection rate was also higher with highest accuracy. Detections were unaffected by the surroundings in method 3 unlike other two methods.

5. CONCLUSION

It can be concluded that this project will help in closing the communication gap. Also, this will remove the need of mediator or translator as this project will serve as the mediator while communicating. This model was able to detect all the signs with high accuracy and the high precision with high speed, which is a better result than all the work mentioned earlier

6. FUTURE SCOPE

Right now model is only able to detect the signs for which it trained for. But we are trying to make this model to learn new signs quickly without all of the training hustle. Also, we are planning to add speech to recognizer so that blind people will also get benefit from this system.

REFERENCES

1. He, Siming. (2019). Research of a Sign Language Translation System Based on Deep Learning. 392-396. 10.1109/AIAM48774.2019.00083.
2. International Conference on Trendz in Information Sciences and Computing (TISC). : 30-35, 2012.
3. sign-language-gesture-recognition by Sarfaraz Masood, Adhyan Srivastava, Harish Chandra Thuwal and Musheer Ahmad. doi:10.1007/978-981-10-7566-7_63
4. A real-time continuous Gesture recognition system for sign language by Rung-Huei Liang, Ming Ouhyoung. DOI:10.1109/AFGR.1998.671007
5. Sign Language Recognition System using Convolutional Neural Network and Computer Vision. IJERTV9IS120029
6. Alisha Menon, Andy Zhou, Senam Tamakloe, Jonathan Ting, Natasha Yamamoto, Yasser Khan, Fres Burghardt, Luca Benini, Ana C Arias, and Jan M. Rabaey. "A wearable biosensing system with in-sensor adaptive machine learning for hand gesture recognition",2021 DOI:10.1038/s41928-020-00510-8
7. "Development of hand gesture detection system using machine learning," Priyanka Parvathy, Kamalraj Subramaniam, G.K.D. Prasanna Venkatesan, P.Karthikaikumar, Justin Varghese, T. Jayasankar, 2020. DOI:10.1007/s12652-020-02314-2
8. Fangtai Guo, Zaixing He, Shuyou Zhang, Xinyue Zhao, Jinhui Fang, Jianrong Tan, "Normalized edge convolutional networks for skeleton-based hand gesture recognition",2021 DOI:10.1016/j.patcog.2021.108044
9. "Vision-based hand gesture identification using deep learning for the interpretation of sign language," Sakshi Sharma and Sukhwinder Singh, 2021 doi.org/10.1016/j.eswa.2021.115657
10. "Hand Gesture Recognition Based on Auto-Landmark Localization and Reweighted Genetic Algorithm for Healthcare Muscle Activities," by Hira Ansar, Ahmad Jalal, Munkhjargal Gochoo, and Kimbum Kim, 2021 DOI:10.3390/su13052961
11. Hyper-parameter tuned light gradient boosting machine utilizing a memetic firefly algorithm for hand gesture detection was developed by Janmenjoy Nayak, Binghnaraj Naik, Pandit Byomkesha Dash, and Alireza Souri, and Vimal Shanmuganathan in 2021. doi.org/10.1016/j.asoc.2021.107478
12. Improve Inter-day Hand Gesture Recognition Via Convolutional Neural Network-based Feature Fusion, Yinfeng Fang, Xuguang Zhang, Dalin Zhou, and Honghai Liu, 2021 DOI:10.1142/S0219843620500255
13. Dynamic hand gesture recognition using a combination of two-level tracker and trajectory-guided features, Shweta Saboo, Joeeta Singha, and Rabul Hussain Laskar, 2021 DOI:10.1007/s00530-021-00811-8
14. Dynamic Hand Gesture Recognition Based on 3D Hand Pose Estimation for Human-Robot Interaction, Qing Gao, Yongquan Chen, Zhaojie Ju, and Yi Liang, 2021 DOI:10.1109/JSEN.2021.3059685
15. A Survey on the Recognition of Sign Language with Efficient Hand Gesture Representation by Priyanka Gaikwad, Kaustubh Trivedi, Mahalaxmi Soma, Komal Bhore, Prof. Richa Agarwal. DOI-ijraset.2022.41963