Sign Language Gesture Recognition using CNN

Rakshith kumar D
dept. of CSE,
Manipal Institute of Technology
Manipal, India
rakshithkademane@gmail.com

Ashwath Rao
dept. of CSE,
Manipal Institute of Technology
Manipal, India
ashwath.rao@manipal.edu

R.Vijaya Arjunan
dept. of CSE,
Manipal Institute of Technology
Manipal, India
vijay.arjun@manipal.edu

***Abstract*—In current society, there is a lack of communication with the deaf. The origin of Sign Language (SL) helped to break down this barrier. Sign language uses visually transmitted sign patterns to communicate meaning to non-sign language users. Normal people are unable to interpret the signs used by the deaf since they are not familiar with their meaning. This system's aim is to find a solution to this issue. This device makes use of camera to record different hand motions. The image is then processed using a variety of techniques. In this study, an enhanced convolutional neural network (CNN) called MobileNetV2 has been used to design the SLR. The primary step is pre-processing the image. Then, an edge detection algorithm is used to determine the edges. The text is displayed once the sign is identified by a template matching algorithm. Since the output is text, it is simple to determine what a particular sign means. Once logged into the system, users can choose to use the sign language translation and recognition features, capture images using OpenCV, and then process them using the trained CNN neural network. Additionally, it makes it easier to interact with the deaf. OpenCV-Python is used in the system's implementation.**

***Index Terms*—Sign Language Recognition, American Sign Language, Deep Learning, CNN**

I. Introduction

There are 15 percent of the world's population with disabilities of various kinds. A total of 466 million people is deaf, more than five percent of the population. In 2050, the population will be approximately 2.7 times larger than it was in 2000, with a projected growth of 500 million people. The speech and hearing abilities of at least 70 million people are impaired. Technologies like gesture and facial recognition have gained significant traction in the sign language field in recent years. Different movements known as gestures are used during communication. Hand or body movements are used.

Gestures used in sign language typically involve visually communicated patterns. Some existing systems for translating sign language taking into consideration hand orientation, hand shape, and hand movement. Every sign in sign language has a specific meaning ascribed to it so that one can easily understand and interpret it.

The people create distinct and unique sign languages depending on their native tongues and geographic locations so no sign language that is widely recognized. Around the world, different sign languages are adopted by people. In India Sign Language uses both the right and left hands to depict a variety of hand gestures. The suggested project focuses on hand position and shape while utilizing American Sign Language. One hand is all that is required for ASL. The system's implementation is therefore made simple. ASL has its own growth path and is independent of all spoken languages.

In a nutshell, the procedure involves utilizing a camera to obtain photographs. then pre-processing the sample, that involves changing the RGB-model image that was acquired to a grayscale image. Afterward, use a clever edge detection algorithm to follow the edges. Finally, this produces the result as text after applying a template-matching technique to find the pattern. This technology eliminates the need for a middle translator by bridging the communication gaps between hearing people and deaf people. It succeeds in its goal of turning motions into language.

II. Related Works

[1].The end of 1990 saw the first recognition of sign language.To recognize it, electrochemical devices were used as a primary method.We investigated parameters such as the position, angle, and angle of the hand using the device. Glove-based systems use this approach.Signers are required to wear a cumbersome device when using this method.Additionally, the recognition system has problems with accuracy and efficiency.With the help of a microphone, the system captured sound and converted it into text using Microsoft's Voice Command and Control Engine. Besides its inefficiency in noisy environments, this system also generated incorrect outputs due to noise included in the input.

[2].Video clips of different gestures of sign language are analyzed.

[3].An audio expression is produced based on the analysis. There is a problem with the frame rate of the animation in this case. The frame rate was decreased manually to make it easier to understand the sign language. It was developed to communicate with DD personnel another system referred to as "Intelligent Assistant".

[4].A glove with different dots on each finger was used to display the signs in this case.A real time photo was captured of the signs using digital input.In order to understand what the sign had been shown, the program examined the dots of the graphics in the image file.Then, the wave files prerecorded as regular language signs are recognized.Using clustering, the

dots were grouped together.Predefined tables were mapped to the results of this clustering.

[8].SLR systems that utilise sensors worn on the body to capture sign language gestures usually comprise of sensor embedded gloves that are worn on the hands. These types of SLR systems are one of the two main approaches to capturing gestures to be classified.

[9].SLR systems designed to use video footage or image data to capture a gesture performed by a user can be further categorised into two methods. The first method involves capturing gesture data using a 3D camera, and the second method involves capturing gesture data using a 2D camera. Therefore the literature to be reviewed in this sub-chapter shall be represented in two further sub-chapters discussing SLR systems which use 3D cameras, and SLR systems which use 2D cameras respectively. An advantage to using a 3D camera over a 2D camera is that the depth capturing capabilities of a 3D camera allow for easier image pre-processing as everything registered as greater in depth than the signer can quickly be removed, solving the issue of complex environmental backgrounds and lighting as seen with 2D cameras.

III. EXPERIMENTAL SETUP

In our research, we used visual studio code to preprocess and train the CNN model. visual studio code is an free IDE for running Python and machine learning code. We built our customized dataset for training. To use the dataset on VS code we need to import all the datasets to VS code. After importing the dataset to VS code we are ready to work on it. To preprocess the dataset and train the model we need some python and machine learning libraries and frameworks. Those libraries and frameworks are mediapipe, OpenCV, Tensorflow, Keras, Sklearn, Numpy, Pandas, etc. We can install those libraries and frameworks using the pip package of python.

IV. DATASET

The system proposed here uses American Sign Language (ASL) to recognize the signs made by the gesture. We made our own samples, though there are many open-source datasets available that we could use, it is recommended to create your own dataset to train your model on as it provides control on number of samples, quality and variety of samples. The Dataset contains numbers labelled from 0 to 9, the alphabets labelled from A to Z and gestures like Hello, Hey, What's Up?, My, Name is, Nice, Meet you, How are you?. This dataset has around 500 samples for every 36 symbols. A few samples are shown in Fig 1. These samples of each symbol cover all the hand shapes and movements. The features are all of right hand. Every sample set is, characterized with its equivalent sign. A unique sign letter corresponds to every sample. The samples are in JPG format with the size of 300x300 pixels.

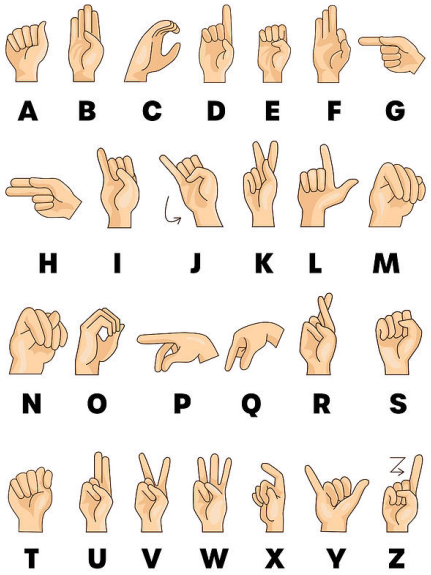

Fig. 1. Alphabets of ASL.

V. METHODOLOGY

A. Gesture Image

We acquire the image using a camera, any type of camera works. The captured images are in the RGB color format. RGB Color Format: RGB represents Red, green, and blue in the RGB color system. Many image-processing techniques use RGB color model as a prerequisite. RGB is a device dependent model. Different systems generate various values. We can normalize the RGB values to remove distortions brought on by light and shadows.

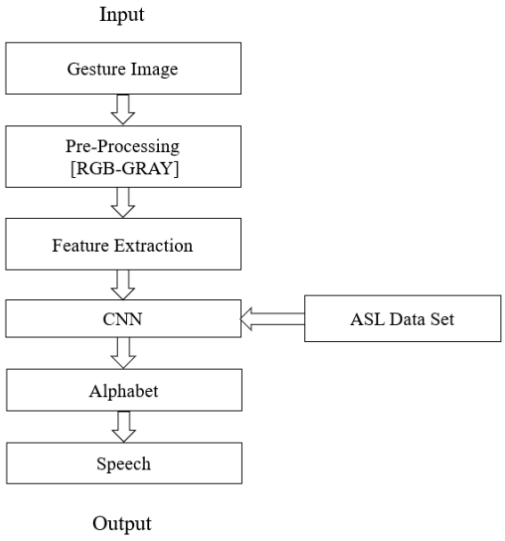

Fig. 2. This is our sign language detection system's main architecture. We can see an overview of our entire approach through this figure.

B. Pre-Processing

Data pre-processing is a crucial part in developing a Machine Learning Model. Here as our data is in RGB format,

so the dimension of our data will be in 3D. Although it's okay to train our model with 3-dimension data, it will be quite computationally expensive for the model to train on. So, we will transform the image from RGB to Grayscale. A grayscale image is nothing but a black white image, this reduces our images to 1D thus making it easier for our model to train. The conversion can be done using OpenCV library's function,

cv2.cvtColor(image, cv2.COLOR$_R GB2GRAY$)

C. Feature Extraction

Here, while we collect our data initially, we collect Hand landmarks. Hand landmarks are the specific points or landmarks detected on a human hand by computer vision algorithms like OpenCV or MediaPipe. The landmarks are necessary for hand gesture recognition, hand pose estimation, and other applications related to hand tracking. MediaPipe provides a more advanced and accurate hand landmark detection module called MediaPipe Hands, which uses a deep learning model to detect and track 21 hand landmarks on each hand. The landmarks include the tip and base of each finger, the center of the palm, and the wrist. The library we are using to collect Hand Landmarks is CVZone. CVZone is a computer vision package that makes it easy to process the image and AI functions. At the core it uses OpenCV and Mediapipe libraries. The hand landmarks in our sample is shown in Fig 5. You can utilize the hand detection module using the below code.,

from cvzone.HandTrackingModule import HandDetector
detector = HandDetector(maxHands=1)

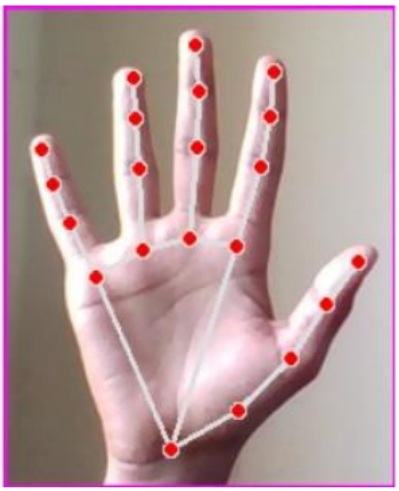

Fig. 3. Hand landmarks in an Image.

The final step is to train our model. In this system we are utilizing a web-based model training called Teachable Machine by Google. Since training a model with large dataset needs huge computation power turning towards a cloud-based service is essential. It has a nice user-friendly UI; Before training the model, we upload the collected samples and label the classes. It provides ability to fine tune our model by changing parameters. We have updated the parameter values of learning rate to 0.0001, epochs to 100 and batch size to 16.

The teachable machine uses transfer learning methodology. It builds our model on pre-trained models like MobileNetV2. MobileNetV2 is a CNN based model of classification that provides easy deployment on mobile devices.

VI. RESULTS

Fig. 4. Predicting Letter 'A' with 96.14% accuracy

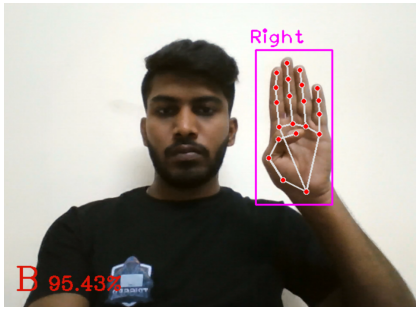

Fig. 5. Predicting Letter 'B' with 95.43% accuracy

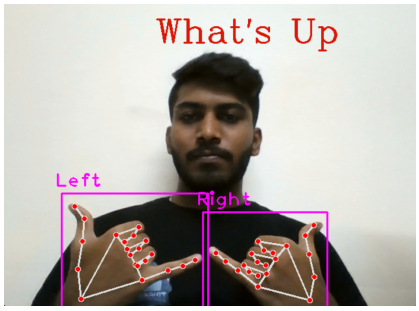

Fig. 6. Predicting Gesture What's up

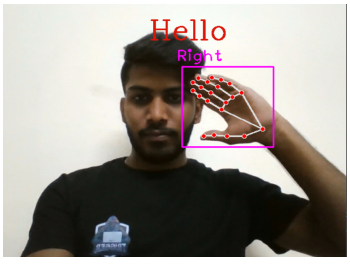

Fig. 7. Predicting Gesture Hello

VII. CONCLUSION

In this paper, A SLR system based on Computer Vision (CV) is designed. The system can translate user input in American Sign Language into the corresponding text by using the CNN neural network to extract characteristics from the ASL data and then giving that information into the MobileNetV2 classifier. For the hearing-impaired groups, experiments assess sign language translation accuracy of 90.45% and recognition accuracy of 93.46% to meet basic demands. There, the procedure entails applying a smoothing algorithm to the image, removing noise and other unimportant data.

REFERENCES

[1] F. M. Rahim, T. E. Mursalin, and N. Sultana, "Intelligent sign language verification system–using image processing, clustering and neural network concepts," International Journal of Engineering Computer Science and Mathematics, vol. 1, no. 1, pp. 43–56, 2010.

[2] D. S. H. Pavel, T. Mustafiz, A. I. Sarkar, and M. Rokonuzzaman, "Geometrical model based hand gesture recognition for interpreting bengali sign language using computer vision," in ICCIT, 2003.

[3] A. Eshaque, T. Hamid, S. Rahman, and M. Rokonuzzaman, "A novel concept of 3d animation based'intelligent assistant'for deaf people: for understanding bengali expressions," in ICCIT, 2002.

[4] S. Rahman, N. Fatema, and M. Rokonuzzaman, "Intelligent assistants for speech impaired people," in ICCIT, 2002.

[5] J. Ubido, J. Huntington, and D. Warburton, "Inequalities in access to healthcare faced by women who are deaf," Health Social Care in the Community, vol. 10, no. 4, pp. 247–253, 2002.

[6] A. Lawson, "United nations convention on the rights of persons with disabilities (crpd)," in International and European Labour Law. Nomos Verlagsgesellschaft mbH Co. KG, 2018, pp. 455–461.

[7] H. Haualand and C. Allen, Deaf people and human rights. World Federation of the Deaf, 2009.

[8] Shanableh T, Assaleh K, Al-Rousan M (2007) Spatio-Temporal Feature Extraction Techniques for Isolated Gesture Recognition in Arabic Sign Language. IEEE Trans on Cybernetics, 37(3), pp.641-650.

[9] Galicia R, Carranza O. et al (2015) Mexican Sign Language Recognition using Movement Sensor. In: IEEE. 2015 IEEE 24th Int Symposium on Industrial Electronics. Buzios, 3-5 June. pp.573-578.

[10] Setiawardhana Hakkun RY, Baharuddin A (2015) Sign Language Learning Based on Android for Deaf and Speech Impaired People. In: IEEE. 2015 Int Electronics Symposium. Surabaya, pp.114-117.

[11] Soodtoetong N, Gedkhaw E (2018) The Efficiency of Sign Language Recognition using 3D Convolutional Neural Networks. In: IEEE. 2018 15th Int Conf on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology. Chiang Rai, 70-73.

[12] Chu TS, Chua AY, Secco EL, A Wearable MYO Gesture Armband Controlling Sphero BB-8 Robot, HighTech and Innovation Journal, 1(4), 179-186, http://dx.doi.org/10.28991/HIJ-2020-01-04-05

[13] Sodhi P, Awasthi N et al (2018) Introduction to Machine Learning and its Basic Application in Python. Proceedings of 10th Int Conf on Digital Strategies for Organizational Success, India, 5-7 January. pp.1-22.

[14] Kanwal K, Abdullah S et al (2014) Assistive Glove for Pakistani Sign Language Translation. In: IEEE. 17th IEEE International Multi Topic Conference 2014. Karachi, 8-10 December, pp.173-176.

[15] Pigou L, Dielemna S et al (2015) Sign Language Recognition Using Convolutional Neural Networks. Computer Vision - ECCV 2014 Workshop. Zurich, Switzerland, 6-7 September. Springer, pp.572-578.