Sign To Text Conversion- Helping Aid

Vatsal Patel*, Maahi Patel

Devang Patel Institute of Advance Technology and Research, Charusat University, Gujarat, India

ABSTRACT

The ancient way of sign language is most natural forms of communication. The recognition of sign is place a key role in research field. The development and improvement on this kind of work need more and more new techniques to analyze the accurate results. Many people don't know it and interpreters are hard to come by, we developed a real-time technique for finger spelling-based American Sign Language using neural networks. In our technique, the hand is first sent through a filter, and then it is passed through a classifier, which analyses the class of hand movements. For each alphabet the proposed model has a 96 percent accuracy rate. This model mainly implemented for Dumb and Deaf people for communication.

Keywords: Sign recognition, CNN, Text conversion, American Sign Language

I. INTRODUCTION

Around the world, there are over 300 distinct sign languages in use. They differ from country to country. Even in nations where some language is spoken, sign language can have a variety of local accents, resulting in slight differences in how users use and comprehend signs. There isn't a sign language that everyone understands. In different countries various sign languages are employed. British Sign Language, for example, is not the same as American Sign Language, and Americans who speak ASL may not comprehend BSL. ASL characteristics are included into the sign languages of some countries.[11][12]

Sign language enhances and improves kid's mental functions, resulting in improved complex and problem-solving skills, thinking skills, cognitive flexibility, academic accomplishment, listening skills, and more. Diversity, literacy, and other cognitive advantages are also promoted. There are many ways for Deaf, mute people to communicate their messages, including notes, sign language, aid pages, lip reading, books with letters, and gestures. Although these pathways, Deaf-mutes and normal individuals have numerous communication issues. [6][8]
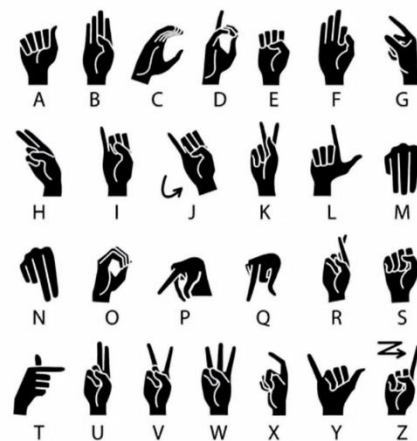

Figure 1. Signs of Each Alphabet

The purpose of this project is to create a functional genuine eyesight American gesture recognition system for Deaf & Mute people who use the ASL alphabet letters. Using the open cv package, we constructed our own data collection. Because each neuron connection delivers data to the next neuron, we employed ANN. The proposed model is built as a combination of two algorithm layers that anticipate the person's final sign. Apply the gaussian blur filter and threshold to the image collected with OpenCV in the first layer to produce the processed image during feature extraction. The CNN prediction model receives the updated image, which is then shown and used to construct the word. Using our method's algorithm layer one and a combination of layers 1 and 2, we were able to achieve excellent results.

The categories for this paper's description are shown below. We analysed a few published studies on sign language to text conversion for Dumb and Deaf people in part II. Several algorithms are combined in section III, with two layers of algorithms checking and forecasting symbols that are more similar to one another. Section V presents a discussion of the observed outcomes and findings, followed by a conclusion.

II. RELATED WORK

Mahesh Kumar N B [1]. In the field of sign language recognition, many novel techniques have been created. The authors demonstrate sign language recognition of twenty-six hand movements in the Indian language using Matlab in this work. Pre-processing, feature extraction, hand segmentation, sign to text, and sign recognition were all used. The LDA algorithm is utilised for action recognition, and the recognised action is transformed into text and audio format, as well as segmentation, using image processing.

Kohsheen Tiku, Jayshree Maloo, Aishwarya Ramesh,Indra R[2].The authors of this paper experimented with various techniques and analysed the results of various methods for converting sign language to speech format. Following the analysis, they developed an Android application to convert real-time American Sign Language signs to speech.

Rajesh Singh, Satyam Shekhar, Shashank Shaurya, Shivang Kumar, Dr. Rekha. K.S [3].Many studies are being conducted in this area. The authors of this paper identified 28 sign languages in ASL, including backspace and whitespace hand movements. They provided five techniques, including caching bottlenecks, setting up the model, and adding the FC layer to the method they presented. They also trained the model and presented the graphs and labels that had been trained.

P. Karthick, N. Prathiba, V. B. Rekha, S. Thanalaxmi [4]. The authors created a leap device that tracks data generated by a leap motion controller, such as [point, reach, grab, wave]. They used DTW in conjunction with the IS algorithm to turn hand actions into flawless text. They introduced a jump device with built-in cameras and two infrared sensors to detect hand motions. They devised an IS algorithm that can alter dynamically depending on the current environment.

Bindu V. and Anuja V. Nair [5]. The authors of this research have identified a number of technologies that are intended to replace sign language translators. They also looked at a variety of image processing methods and AI that has recently been introduced in this field. Because the ISL is double-handed and more sophisticated than single-handed ASL, several of the authors found it difficult to use. Many Indian academics have recently started researching ISL in order to develop automatic ISL recognition systems. Working on ISL entails mostly three steps: Pre-processing, classification, and feature extraction for

sign languages. For sign language recognition, they used ANN, SVM, and HMM.

Tokuda, Y. Nankaku, T. Toda, H. Zen, J. Yamagishi, K. Oura [6]. The author of this paper proposed a new approach to speech synthesis based on the HMM (Hidden Markov model), which has recently been shown to be very effective in speech processing. The flexibility of this strategy in changing speaker identities, speaking styles, and emotions is the fundamental reason for employing it. They also talked about how the HMM-based technique compares to the more traditional unit selection methods that have been utilised for decades.

III. METHODS AND MATERIAL

We tried to discover ready-made datasets for the project, but we couldn't find any in the form of raw pictures that met our specifications. We could only locate data in the form of RGB values. As a result, we planned to construct our own data collection. The following are the steps we used to build our data set. To create our dataset, we utilized the OpenCV package. To begin, we took around 800 photos of each ASL sign for training reasons and approximately 200 images of each sign for testing purposes. First, we grab each image displayed by our computer's webcam. We designate an ROI in each image, which is represented by a blue-bordered square. We retrieve our ROI, which is RGB, from this entire image and transform it into an image pixel, as seen below. Finally, we apply our Gaussian blur filter to our picture, which aids in the extraction of numerous characteristics. The Block diagram of generating the dataset Fig2.
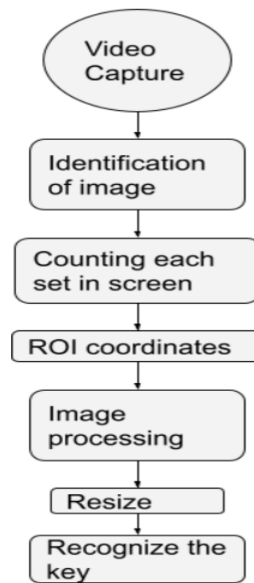

Figure 2. Block Diagram of the proposed method

The data that has been collected is now ready to be used to create the model. A convolution neural network and a few models are used to build the model. To begin, the layers are fully linked to one another and the model is being compiled. Image processing is used to simplify and characterize an image by blurring the background, adjusting the adaptive threshold, and changing the colour of the image. The information is separated into two groups: training and testing. Image data generators are used in pre-processing to zoom, rescale, and relocate images, among other things. To save the model, the training and testing sets are loaded into the generator.
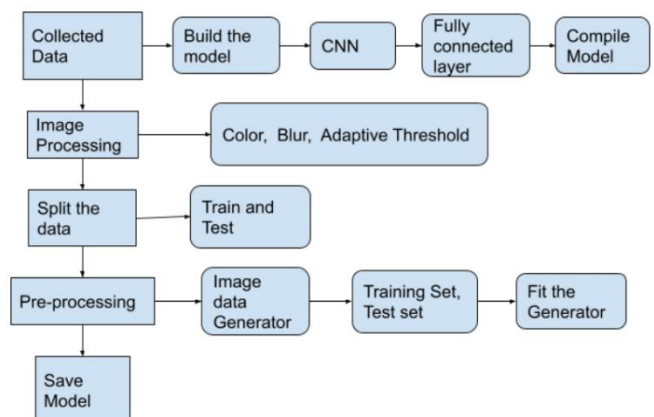

Figure 3. Block Diagram of the training model

An ANN is a network of neurons that mimics the components of the human brain. Every neuron's connection sends data to the next neuron. Inputs are delivered into the first layer of neurons, which analyse them before passing them on to the buried layers. Data is transmitted to the ultimate hidden layers after processing via many levels of the convolution layer.[13]

In contrast to conventional Neural Networks, the neurons in CNN layers are organized in three dimensions: width, height, and depth. Instead of connecting all of the neurons in a layer in a completely linked way, the neurons in a layer will only be linked to a limited area of the layer preceding it. Furthermore, because we will compress the whole picture into a single vector of class scores towards the conclusion of the Deep CNN, the final feature layer will have dimensions.[14]

The suggested model is implemented as a combination of two algorithm layers that anticipate the final sign from the user. In the first layer, to obtain the processed image during extracting features, apply the gaussian blur filter and threshold to the image captured using OpenCV [15]. This modified image is sent to the CNN prediction model, but if a character is identified for more than 50 frames, it is displayed and used to create the word. The empty symbol represents the gap between the words. In the second layer, we identify several sets of symbols that have comparable effects when detected. We then use classifiers designed specifically for those sets to differentiate between them. The algorithm of the proposed model is described as below Figure4.


```
Algorithm: Proposed model using CNN

Input: Raw input
Output: Recognition of sign and convert into text format
Initialization:
Construction of CNN model
Pre-processing
for files in folder
        for sign name in file
            for extraction of img
                If img is not in train path and mg is not in test path
                    Reset the format
                N = 10^17
                for file in files
                    Save the path
                    actual_path1=path1+"/"+"train/"+dirname+"/"+file
                    actual_path2=path1+"/"+"test/"+dirname+"/"+file

                    Read img
                    If i =0 < num: c1= c1+1 and img write with path of train and test
                end
            end
        end
end
Test the signs with Constructed model
Convert signs into Text
```

Figure 4. Pre-processing Algorithm

IV.   RESULTS AND DISCUSSION

To eliminate unwanted noise, we transform our RGB input pictures to grayscale and apply gaussian blur. To remove our hand from the backdrop, we use an adaptive threshold and scale our pictures to 128 by 128 pixels. We send the pre-processed input pictures to our model for training and testing after doing all of the preceding procedures. The prediction layer calculates the likelihood of the picture falling into one of the classifications. As a result, the output is normalized between 0 and 1, and the total of each class's values equals 1. Using the soft-max function, we were able to do this.
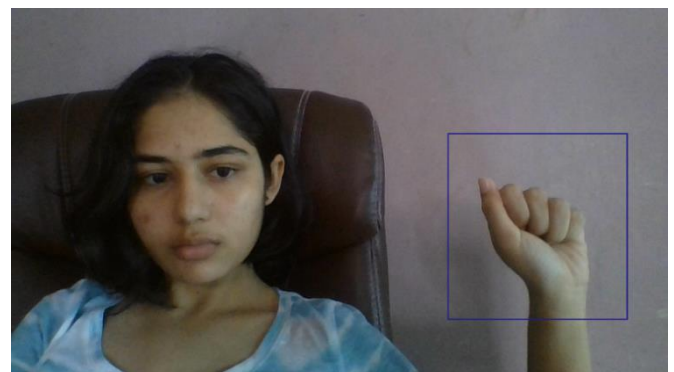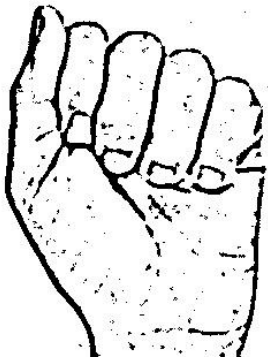

Figure 5. Region of Interest

Figure 6. Image Processing

The prediction layer's output will initially be a little off from the real value. To improve it, we used labelled data to train the networks. Cross-entropy is a performance metric that is used in classification. It is a continuous function that is positive when the value is not the same as the labelled value and zeroes when the value is the same as the labelled value. As a result, we maximized the cross-entropy by bringing it as near to zero as possible. We modify the weights of our neural networks at our network layer to do this. The cross-entropy may be calculated using TensorFlow's built-in function. We used Gradient Descent to improve the cross-entropy function after discovering it. The best gradient descent optimizer is named Adam Optimizer.

Figure 7. Result

The proposed model achieved 96% using algorithm layer one, and a 98.0 percent accuracy using a combination of layer 1 and layer 2 of our method, which is better than most current research publications on American sign language. The majority

of the research articles focus on employing Kinect-like devices to detect hands. In [7], they use convolutional neural networks and Kinect to create a recognition system for Dutch sign language with a 2.5 percent mistake rate. [8] uses a hidden Markov model classifier with a vocabulary of 30 words to create a recognition model with a 10.90 percent error rate. In [9], they obtain an average accuracy of 86 percent in Japanese sign language for 41 static motions. With a device that measures, Map [10] obtained an accuracy of 99.99 % for existing signers and 83.58 % and 85.49 % for new signers.

Their recognition system was likewise based on CNN. One thing to keep in mind is that our model does not employ a background subtraction technique, although some of the other models do. As a result, the accuracy of background subtraction may vary depending on how we apply it in our project. The majority of the above 21 projects, on the other hand, utilize Kinect devices, but our main goal was to build a project that could be used using easily available resources. Because a sensor like Kinect is not only not widely available but also costly for most of the audience, our approach, which uses a standard laptop camera, is a huge bonus. The confusion matrices for our results are shown below.

V. CONCLUSION

This study describes the development of a workable authentic vision-based American gesture recognition for Deaf & Mute individuals using asl alphabet letters. On our dataset, we attained an ultimate accuracy of 98.0 percent. We can enhance our predictions after installing 2 layers of algorithms that check and forecast symbols that are more similar to one another. This method allows us to identify virtually any signs as long as they are correctly displayed, there is no noise in the surrounding, and the illumination is appropriate.

VI. REFERENCES

[1] NB, M. K. (2018). Conversion of sign language into text. *International Journal of Applied Engineering Research*, *13*(9), 7154-7161.

[2] K. Tiku, J. Maloo, A. Ramesh and I. R., "Real-time Conversion of Sign Language to Text and Speech," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), 2020, pp. 346-351, doi: 10.1109/ICIRCA48905.2020.9182877.

[3] Huang, J., Zhou, W., Li, H., & Li, W. (2015, June). Sign language recognition using 3d convolutional neural networks. In *2015 IEEE international conference on multimedia and expo (ICME)* (pp. 1-6). IEEE.

[4] Karthick, P., Prathiba, N., Rekha, V. B., & Thanalaxmi, S. (2014). Transforming Indian sign language into text using leap motion. *International Journal of Innovative Research in Science, Engineering and Technology*, *3*(4), 5.

[5] Anuja V.Nair, Bindu.V, "A Review on Indian Sign Language Recognition", International journal of computer applications, Vol. 73, pp: 22, (2013).

[6] Tokuda, K.; Nankaku, Y.; Toda, T.; Zen, H.; Yamagishi, J.; Oura, K., "Speech Synthesis Based on Hidden Markov Models", in Proceedings of the IEEE, vol.101, no.5, pp.1234-1252, (2013).

[7] Archana S. Ghotkar, Rucha Khatal, Sanjana Khupase, Surbhi Asati & Mithila Hadap, "Hand Gesture Recognition for Indian Sign Language", IEEE International Conference on Computer Communication and Informatics (lCCCI), pp: 1-4, (2012).

[8] Pigou L., Dieleman S., Kindermans PJ., Schrauwen B. (2015) Sign Language Recognition Using Convolutional Neural Networks. In: Agapito L., Bronstein M., Rother C. (eds) Computer Vision - ECCV 2014 Workshops. ECCV 2014. Lecture Notes in Computer Science, vol 8925. Springer, Cham

[9] Zaki, M.M., Shaheen, S.I.: Sign language recognition using a combination of new vision-based features. Pattern Recognition Letters 32(4), 572–577 (2011)

[10] N. Mukai, N. Harada and Y. Chang, "Japanese Fingerspelling Recognition Based on Classification Tree and Machine Learning," 2017 Nicograph International (NicoInt), Kyoto, Japan, 2017, pp. 19-24. doi:10.1109/NICOInt.2017.9

[11] Byeongkeun Kang, Subarna Tripathi, Truong Q. Nguyen" Real-time sign language fingerspelling recognition using convolutional neural networks from depth map" 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)

[12] Ruize Xu, Shengli Zhou, Li, W.J, "MEMS Accelerometer Based Nonspecific-User Hand Gesture Recognition", IEEE Sensors Journal, vol.12, no.5, pp.1166-1173, (2012).

[13] Rajam, P. Subha and Dr G Bala krishnan, "Real Time Indian Sign Language Recognition System to aid Deaf and Dumb people", 13thInternational Conference on Communication Technology (ICCT), pp. 737-742, (2011).

[14] Prof. Rajeshri Rahul Itkarkar, "A Study of Vision Based Hand Gesture Recognition for Human Machine Interaction", International Journal of Innovative Research in Advanced Engineering, Vol. 1, pp:12, (2014).

[15] Pravin R Futane, Rajiv V Dharaskar, "Hasta Mudra an interpretatoin of Indian sign hand gestures", IEEE 3 rd International Conference on Electronics Computer Technology, Vol.2, pp:377-380, (2011).

[16] Meenakshi Panwar, "Hand Gesture Recognition based on Shape Parameters" International Conference on Computing, Communication and Application (ICCCA), pp: I-6, IEEE, (2012).

[17] Baker-Shenk, C. L., & Cokely, D. (1991). *American Sign Language: A teacher's resource text on grammar and culture*. Gallaudet University Press.

[18] Yegnanarayana, B. (2009). *Artificial neural networks*. PHI Learning Pvt. Ltd.

[19] Liu, T., Fang, S., Zhao, Y., Wang, P., & Zhang, J. (2015). Implementation of training convolutional neural networks. *arXiv preprint arXiv:1506.01195*.

[20] Bradski, G., & Kaehler, A. (2008). *Learning OpenCV: Computer vision with the OpenCV library*. " O'Reilly Media, Inc.".

[21] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)* (pp. 265-283).

Cite this article as :