DOI: 10.17148/IJARCCE.2022.11719

HAND GESTURE RECONIGITION for PHYSICALLY CHALLENGED

Sahana Ramesh[1], Mohan Kumar H P [2]

Research Scholar, Dept. of MCA, P.E.S College of Engineering, Mandya, India[1]

Professor, Dept. of MCA, P.E.S College of Engineering, Mandya, India[2]

Abstract: Hand gestures are an effective form of communication, particularly when we are speaking to others who cannot comprehend our signing. It's also a crucial component of human-computer interaction. To ensure that listeners comprehend what speakers are attempting to say, understanding hand gesture is crucial. Speaking will be helpful for the deaf and the dumb, and the speaking mouth will be helpful for the stupid. Convolutional Neutral Networks (CNN) are frequently used to categorise photographs of hand gestures. Voice recognition is complemented by the conversion of hand movements into text picture manipulation. It offers more accuracy.

Keywords: CNN, Image Processing, Deep Learning, Feature extraction, Vision based system.

I. INTRODUCTION

Language has undoubtedly been an aspect of human interaction since the beginning of civilisation. People use it as a method of communication in order to express themselves and get knowledge of the outside world. Because it is so engrained in our daily lives, we frequently take it for granted and fail to recognise its significance. In our fast-changing environment, it is a sad reality that people with hearing loss are regularly disregarded and ostracised. They devote a lot of time and effort to persuading people of their viewpoints and views. The most efficient means of communication for those who have difficulty hearing or speaking is unquestionably sign language. For them, it is made possible and simpler to get along with other people. However, only the development of sign language is insufficient. This advantage has a lot of restrictions. The sign motions typically become confusing and difficult to grasp for someone learning sign language for the first time or who is learning it in a foreign language. Although sign language is a means of communication for the deaf, it has no meaning when employed by someone who is not deaf increasing the communication gap. This long-standing communication gap may now be closed thanks to the development of many automated approaches for the detection of sign motions. A convolutional neural network (CNN) that has been trained using a model is used in the general approach to recognise indicators. The model may extract features from each of its numerous layers by

Identifying specific features indicates if a new image corresponds to the training dataset. With this strategy, we suggest a useful and clever application for hand gesture recognition.

II. RELATED WORK

According to a survey of the literature on the subject, numerous methods and algorithms have been investigated to address the problem of sign identification in films and photos.

"Sign Language Recognition Using Machine Learning by Hemlata Dakhore, Manali Landge, Shivani Patil, Tanushree Patil, Shrutika Zyate, Ashwini Moon, Raveena Lade International Journal of All Research Education and Scientific Methods June -2021- To detect sign language motions in this instance, a unique CNN model is utilised. In order to train the model to recognise the gesture, they use the American Sign Language Dataset from MNIST. Features of several augmented gestures are included in the dataset.[1]"

"Convolutional Neural Network Hand Gesture Recognition for American Sign Language by Shruti Chavan, Xinrui Yu and Jafar Saniie Publication: Embedded Computing and Signal Processing Research Laboratory (http://ecasp.ece.iit.edu/) Department of Electrical and Computer Engineering Illinois Institute of Technology, Chicago, IL, U.S.A. March-2021- This research uses a model to extract temporal and spatial characteristics from video sequences. To recognise spatial information, Inception, a CNN (Convolutional Neural Network), is then employed. Temporal characteristics are trained using recurrent neural networks, or RNNs. American Sign Language Dataset is the dataset in question.[2]"

"Sign Language Recognition Using Deep Learning and Computer Vision by R.S. Sabeenian, S. Sai Bharathwaj, M. Mohamed Aadhil - Journal of Advanced Research in Dynamical and Control Systems. May 2020- Here a custom CNN model is used to recognize gestures in sign language. They approach the American Sign Language Dataset from MNIST to train the model to identify the gesture. The dataset contains the features of different augmented gestures. [3]"

"Research of a Sign Language Translation System Based on Deep Learning by He, Siming, (2019) – The goal of the project is to employ an American Sign Language (ASL) dataset and a neural network composed of the R-CNN algorithm with the LSTM framework and 3D CNN for feature extraction.[4]"

"Static Sign Language Recognition Using Deep Learning by Lean Karlo S. Tolentino, Ronnie O. Serfa Juan, August C. Thio-ac, Maria Abigail B. Pamahoy, Joni Rose R. Forteza, and Xavier Jet O. Garcia - International Journal of Machine Learning and Computing December 2019- The proposed study aims to develop a system that will recognize static sign gestures and convert them into corresponding words. A vision-based approach using a web camera is introduced to obtain the data from the signer and can be used offline.[5]"

"American Sign Language Recognition using Deep Learning and Computer Vision by Kshitij Bantupalli and Ying Xie - IEEE International Conference on Big Data (Big Data) April 2018-In this paper, the model takes video sequences and extracts temporal and spatial features from them. Then Inception, a CNN (Convolutional Neural Network) is used for recognizing spatial features. RNN (Recurrent Neural Network) is used to train on temporal features. The dataset used is the American Sign Language Dataset.[6]"

"Sign Language Recognition using 3D convolutional neural networks by Huang, J., Zhou, W., & Li, H, (2015). IEEE International Conference on Multimedia and Expo (ICME) –In this paper, 3D convolutional neural network (CNN) which extracts discriminative spatial-temporal features from raw stream automatically without any prior knowledge, avoiding designing features. To boost the performance, multi-channels of video streams, including colour information, depth clue, and body joint positions, are used as input to the 3D CNN in order to integrate colour, depth and trajectory information. The model is on a real dataset collected with Microsoft Kinect and demonstrate its effectiveness over the traditional approaches based on hand-crafted features.[7]"

"ChaLearn Looking at People Challenge 2014: Dataset and Results by Escalera, S., Baró, X., Gonzàlez, J., Bautista, M., Madadi, M., Reyes, M., Guyon, I (2014). Workshop at the European Conference on Computer Vision (pp. 459-473). Springer Cham. - Three different tracks made up the competition: multi-modal gesture recognition from RGB-Depth sequences, action and interaction recognition from RGB data sequences, and human pose recovery from RGB data. The overlapping Jaccard index was used as the assessment metric to conduct user-independent recognition in continuous picture sequences for all tracks.[8]"

"IMAGE BASED SIGN LANGUAGE RECOGNITION SYSTEM FOR SINHALA SIGN LANGUAGE by Herath, H.C.M. & W.A.L.V.Kumari, & Senevirathne, W.A.P.B & Dissanayake, Maheshi, (2013). - The objective of this project is to create a device that will enable a hearing-impaired individual to communicate with a person who is not familiar with sign language languages. The method for creating a real-time Sinhala sign language recognition program me based on image processing is presented in this research as being low-cost.[9]"

"International Conference on Trendz in Information Sciences and Computing (TISC) – The MINIST dataset is used to extract the feature from the hand gestures and techniques concerning sign language for visually impaired persons.[10]"

"Multimedia Tools and Springer Hand gesture recognition from depth and infrared Kinect data for CAVE applications interaction by DQ Leite, JC Duarte, LP Neves, 2017 - This paper presents a real-time framework that combines depth data and infrared laser speckle pattern (ILSP) images, captured from a Kinect device, for static hand gesture recognition to interact with CAVE applications. At the startup of the system, background removal and hand position detection are performed using only the depth map. After that, tracking is started using the hand positions of the previous frames in order to seek for the hand centroid of the current one.[11]"

"Gesture recognition based on HMM-FNN model using a Kinect, J. Multimodal User Interfaces, by X. L. Guo and T. T. Yang 11(1), 1–7(2017) – In this study, the body depth picture is obtained using an IOT device that measures body sensation, and the gestures depth image is segmented using the threshold segmentation method Common separation between the hand and the body Next, the HMM-FNN model, which incorporates the Fuzzy Neural Network (FNN) and Hidden Markov Model (HMM) are utilised for dynamic gesture identification. [12]"

"A Vision Based Recognition of Indian Sign Language Alphabets and Numerals Using B-Spline Approximation by M. Geetha and U. C. Manjusha, International Journal on Computer Science and Engineering (IJCSE) – Indian Sign Language is used as a dataset in this study. The alphabets and digits of Indian Sign Language are recognised using a technique called Bag of Visual Words Model (BOVW). It accomplishes this while removing the background by using the algorithms CNN, SVM, and GUI.[13]"

"Sign Language Recognition Using Convolutional Neural Networks by Pigou L, Dieleman S, Kindermans PJ, Schrauwen B, (2015) In: Agapito L., Bronstein M., Rother C. (eds) Computer Vision - ECCV 2014 Workshops. ECCV 2014, Lecture Notes in Computer Science - In this method, the backdrop of the image is made black using the HSV technique, and the colours are shown in grayscale. American Sign Language is also utilised, however there are only 10 datasets available.[14]"

DOI: 10.17148/IJARCCE.2022.11719

III. METHODOLOGY

Camera was used to record gestures, and speaker was used to transfer speech. Software components that recognise hand gestures interpret their position, size, etc. the entire software model is built in the Python environment. The camera method in this instance uses OpenCV. A library of programming functions with a focus on real-time computer vision is called OpenCV (Open-Source Computer Vision). It is an image processing library, to put it simply. The majority of its operations are devoted to images.

A. Dataset:

We first created a custom dataset that is relevant to American sign language and is made up of a combination of various signs and their corresponding labels in order to train the suggested model. The same dataset was used to train the model as a result. There are 10,000 photos in this collection of data. Each graphic features a hand forming an ASL letter (with some variation). Based on the following ASL Alphabet data, the goal of this data set is to serve as a sort of validation set to assess how effective the pre-processing and the model are. The W207 Applied Machine Learning course at UC Berkeley's master of information and data science programme (MIDS) is using this data set for a work.

B. Normalization:

Once the image is rendered, the calculating phase begins. Software analyses the entire image to determine the HSV, threshold, and contour of the image. A contour, which represents the shapes of objects found in an image, is a closed curve uniting all the continuous points of some colour or intensity. Shape analysis, object detection, and object recognition can all benefit from the usage of contour detection. We locate the locations where the intensity of the colours dramatically varies when we perform edge detection, and we then simply turn on those pixels. Contextual groupings of points and segments, on the other hand, are abstract representations of the forms of the objects in an image. In our software, we can therefore alter contours by counting the number of contours, employing them to crop things from an image (image segmentation), classifying the shapes of objects, and much more. For the model to accurately forecast results and help us build a stronger model, we need a specific set of photos and their related labels.

C. System Architecture:

A developed system or systems architecture is the conceptual model that describes the composition, operation, and other viewpoints of a system. An official description and representation of a system, created to make it easier to analyse its structures and behaviours, is known as a system architectural description. It offers a visual summary, which makes it easier to communicate ideas and crucial concepts. In system architecture, basic lines and shapes are used to represent components and relationships. The architecture describes sign language; the webcam will first display sign language and record movement; picture pre-processing will then be carried out; the feature is extracted; and the sign language will be recognised.

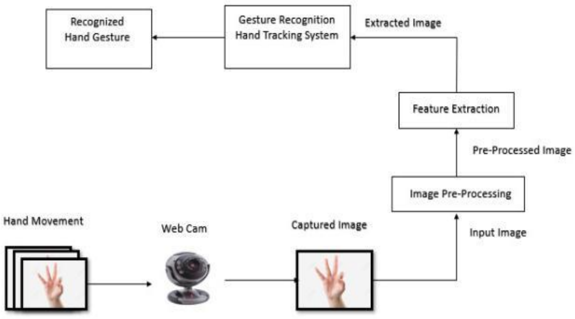


Figure 1 System architecture for Hand Gesture

D. Data Pre-Processing:

The background blurring and edge detection processes, which are used to identify the hands in the image, are some common stages that were addressed to execute data analysis. Certain pixels must be mapped in order to validate the best data needed to do a particular analysis on the data. This procedure is crucial for better model training since it purges the image of all undesirable information. Convolution is a specialised linear procedure used to extract important features from data. By reducing the covariance shift, batch normalisation is utilised to improve the neural network's stability. By subtracting the batch mean and dividing by the batch, it normalises the output of an earlier activation layer average deviation The rectified linear activation function is a piecewise linear function that outputs zero otherwise and the input directly if the input is negative. The ReLU function may be described using equations (1) and (2). ReLU has a constant gradient for the positive input.

$\mathrm{ReLu}(x) = \mathrm{Max}(0, x),$ (1)

d ReLu(x) ={1, if x > otherwise} (2)

dx

E. Convolutional Neural Network:

Convolutional neural networks are feed-forward neural networks that process data in a grid-like layout to analyse visual images. A ConvNet is another name for it. A convolutional neural network is used to locate and classify objects in a picture.

Layer of a convolutional neural network

A convolution neural network's several hidden layers help extract data from images. The four main tiers of CNN are as follows:

1. A convolutional layer
2. The initial ReLU layer.
3. A pooling layer
4. Completely linked layer

IV. IMPLEMENTATION

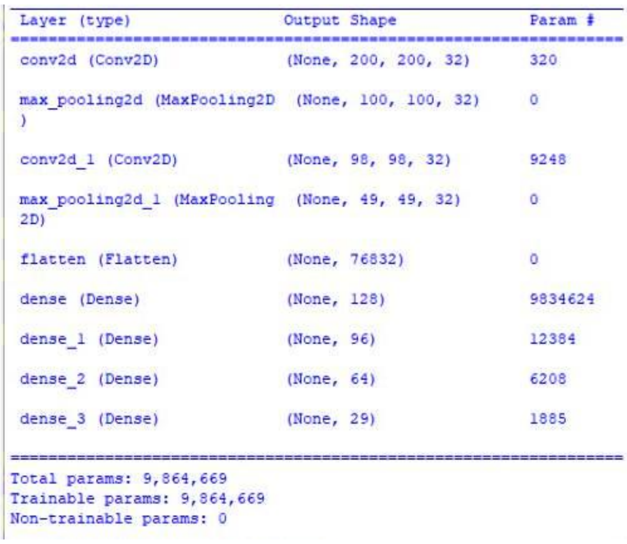
Data gathering is the initial stage of the proposed system. In numerous studies, sensors or cameras have been used to record hand movements. We use the web camera to capture hand motions for our system. The backgrounds in the images are recognised and eliminated using a number of post-production stages utilising the colour extraction method HSV (Hue, Saturation, Value). Segmentation is then used to determine the location of the skin tone. The images are given a mask using morphological techniques, and an elliptical kernel is used to carry out a series of dilation and erosion processes. Since OpenCV scales all of the photographs to the same size, it is difficult to distinguish between photos taken during different activities. In our dataset, 2000 images of American sign gestures are included 400 are used for testing, while 1600 are used for training. It is an 80:20 split. An extraction of the binary pixels from every frame is utilised to train and classify a convolutional neural network. After the model has been assessed, the system will be able to predict the alphabets. A CNN model is used to predict hand motions and extract data from the frames. The main use of this multi-layered feedforward neural network is image recognition. The convolution layers that make up the CNN architecture each have a pooling layer, an activation function, and an optional batch normalisation. There are also a few layers that are completely integrated. One of the images shrinks while it's being transmitted across the network. Max Pooling is the main problem. The chance for each class is predicted by the top layer. The Convolutional Neutral Network machine learning algorithm and the Open CV library for image processing were utilised in the development of the suggested system. The suggested task For those who have trouble hearing, hand gestures are an effective form of communication. The deaf and dumb employ a hand gesture language that is difficult for anyone who are not familiar with it to understand. Therefore, it is necessary to develop a technology that can translate motions into speech. The major objective of this effort is to develop a system that can automatically translate sign language gestures into spoken speech. It is a method known as artificial speaking organ for the mentally impaired. The algorithm converts the identified sign into text that makes sense. Speech synthesis will be applied to this text. As a result, the system can adapt to the extensive variance among sign languages.

V. RESULT AND ANALYSIS

In the proposed work, we have suggested a model to predict the most likely characters using minimal threshold selection, utilising OpenCV, package cv2 put-Text. The prediction made by the suggested model is 95% accurate. To attain this accuracy, we established two distinct folders based on the training and test set sizes, which were obtained by building a dataset of 10,000 photos, 390 of which were used for the training and test sets, respectively. Additionally, we identified the characters for which we used those sets of photos to train CNN models. For those photographs that have been enhanced with hand images utilising certain convolution and other linear operations to enlarge the set of images, this was trained with 50 epochs to reach that accuracy. The model is being trained using 50 epochs which signifies 50 times the weights are modified to follow-up the prediction of the model. The formula we employ is Accuracy=TP+TN/TP+TN+FP+FN.

Using the location, the ratio of true positive to true negative, and this, we predict that the accuracy will be 95%. After retraining on those images, we achieve 95.07 percent as a result. All 26 of the English alphabet's letters have been recognised for use in American sign language.

DOI: 10.17148/IJARCCE.2022.11719



```
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 200, 200, 32)      320

 max_pooling2d (MaxPooling2D  (None, 100, 100, 32)     0
 )

 conv2d_1 (Conv2D)           (None, 98, 98, 32)        9248

 max_pooling2d_1 (MaxPooling  (None, 49, 49, 32)       0
 2D)

 flatten (Flatten)           (None, 76832)             0

 dense (Dense)               (None, 128)               9834624

 dense_1 (Dense)             (None, 96)                12384

 dense_2 (Dense)             (None, 64)                6208

 dense_3 (Dense)             (None, 29)                1885

=================================================================
Total params: 9,864,669
Trainable params: 9,864,669
Non-trainable params: 0
```

Table 1 Output of Hand Gesture

Comparison with other models:

- In this work the background can have any clear colour it had overcome the white background in the earlier system.[1].
- In our proposed system we have built a real time application for deaf and dumb people it can be used in any platform it had overcome the embedded system which they are using for specific function and limited speed for smartphones.[2].
- The dataset is taken is American Sign Language (ASL) which is better accurate than the MINST dataset.[3].
- The algorithm used is CNN algorithm which is best for image processing and feature extraction is used has OpenCV which is better than LSTM and RNN algorithm.[4].
- Our dataset is both static and dynamic, it involves only static gesture means skin colour.[5].
- The dataset takes the image for feature extraction and extracted feature will be trained by CNN algorithm. [6].
- The dataset is American Sign Language (ASL) which has plenty of images.[7].
- Here, our model has HSV which used for colour vision based need not to be divided into tracks.[8].
- ASL dataset sign gesture is easier than this model. [9].
- Our work is used for both dumb and deaf people to be easier for them. [10].
- Open CV which used for vision-based images which do all operations related to the images. [11].
- For recognising the alphabet, we will train and test the model by using the CNN algorithm so it become easier. [12].
- Only one algorithm can be used for many operations by using their libraries.[13].
- Here, we have used dataset around 10000 sign languages which gives more accuracy.[14].

VI. CONCLUSION

A system that could only successfully recognise static signs and alphabets has grown to be able to understand dynamic motions that happen in nonstop streams of images. The development of a complete lexicon for sign language recognition systems is currently the main focus of research. A number of academics are developing their own sign language recognition systems using their own databases and a restricted vocabulary. A considerable database that was

DOI: 10.17148/IJARCCE.2022.11719

developed is currently inaccessible for some of the countries using sign language recognition devices. The neural network is one of the more effective methods for pattern recognition and identification systems.
By automatically eliminating the background from the captured frame, the model may be further trained using a dataset so that it can automatically separate the gesture from the frame. enhancement and tuning of the using a model, you can find common words and expressions.

REFERENCES

[1] Sign Language Recognition Using Machine Learning by Hemlata Dakhore, Manali Landge, Shivani Patil, Tanushree Patil, Shrutika Zyate, Ashwini Moon, Raveena Lade, International Journal of All Research Education and Scientific Methods June -2021**.**
[2] Convolutional Neural Network Hand Gesture Recognition for American Sign Language by Shruti Chavan, Xinrui Yu and Jafar Saniie, Embedded Computing and Signal Processing Research Laboratory (http://ecasp.ece.iit.edu/) Department of Electrical and Computer Engineering Illinois Institute of Technology, Chicago, IL, U.S.A.
[3] Sign Language Recognition Using Deep Learning and Computer Vision by R.S. Sabeenian, S. Sai Bharathwaj, M. Mohamed Aadhil Article in Journal of Advanced Research in Dynamical and Control Systems, May 2020.
[4] He, Siming. (2019). Research of a Sign Language Translation System Based on Deep Learning. 392-396. 10.1109/AIAM48774.2019.00083.
[5] Static Sign Language Recognition Using Deep Learning by Lean Karlo S. Tolentino, Ronnie O. Serfa Juan, August C. Thio-ac, Maria Abigail B. Pamahoy, Joni Rose R. Forteza, and Xavier Jet O. Garcia. International Journal of Machine Learning and Computing, Vol. 9, No. 6, December 2019.
[6] American Sign Language Recognition using Deep Learning and Computer Vision by Kshitij Bantupalli and Ying Xie 2018 IEEE International Conference on Big Data (Big Data).
[7] Huang, J., Zhou, W., & Li, H. (2015). Sign Language Recognition using 3D convolutional neural networks. IEEE International Conference on Multimedia and Expo (ICME) (pp. 1-6). Turin: IEEE.
[8] Escalera, S., Baró, X., Gonzàlez, J., Bautista, M., Madadi, M., Reyes, M., . . . Guyon, I. (2014). ChaLearn Looking at People Challenge 2014: Dataset and Results. Workshop at the European Conference on Computer Vision (pp. 459-473). Springer Cham.
[9] Herath, H.C.M. & W.A.L.V.Kumari, & Senevirathne, W.A.P.B & Dissanayake, Maheshi. (2013). IMAGE BASED SIGN LANGUAGE RECOGNITION SYSTEM FOR SINHALA SIGN LANGUAGE.
[10] International Conference on Trendz in Information Sciences and Computing (TISC).: 30-35, 2012.
[11] DQ Leite, JC Duarte, LP Neves, Multimedia Tools and, Springer Hand gesture recognition from depth and infrared Kinect data for CAVE applications interaction 2017
[12] X. L. Guo and T. T. Yang, “Gesture recognition based on HMM-FNN model using a Kinect,” J. Multimodal User Interfaces, 11 (1), 1 –7 (2017).
[13] M. Geetha and U. C. Manjusha, “A Vision Based Recognition of Indian Sign Language Alphabets and Numerals Using B-Spline Approximation”, International Journal on Computer Science and Engineering (IJCSE).
[14] 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.