e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:06/June-2024 Impact Factor- 7.868 www.irjmets.com

SIGN LANGUAGE DETECTION USING DEEP LEARNING

Himesh Mishra[*1], Rahul Rathour[*2], Mohd Abuzar Suhail[*3], Daya Shankar Srivastava[*4], Mr. Arjun Singh[*5]

[*1,2,3,4,5,6]Greater Noida Institute Of Technology, Greater Noida, India.

DOI : https://www.doi.org/10.56726/IRJMETS58981

ABSTRACT

According to the 2011 Census, In India, out of the total population of 121 crores, approximately 2.68 Crore humans are 'Disabled' (2.21 % of the whole population)). Sign Language serves as a means for these people with special needs to communicate with others, but it is not a simple task. This barrier to communication has been addressed by researchers for years. The goal of this study is to demonstrate the MobileNets model's experimental performance on the TensorFlow platform when training the Sign language. Language Recognition Model, which can drastically reduce the amount of time it takes to learn a new language. Classification of Sign Language motions in terms of time and space Developing a portable solution for a real-time application. The Mobilenet V2 Model was trained for this purpose and an Accuracy of 70% was obtained. The four different sequential combinations of LSTM and GRU (as there are two layers of LSTM and two layers of GRU) were used with our own dataset, IISL2020. The proposed model, consisting of a single layer of LSTM followed by GRU, achieves around 97% accuracy over 11 different signs. This method may help persons who are unaware of sign language to communicate with persons whose speech or hearing is impaired.

I. INTRODUCTION

Since the beginning of Evolution, Humans have kept evolving and adapting to their available surroundings. Senses have developed to a major extent. But unfortunately, some people are born special. They are called special because they lack the ability to use all their five

senses simultaneously. According to WHO, About 6.3% i.e. About 63 million people suffer from an auditory loss in India. Research is still going on in this context.

According to Census 2011 statistics, India has a population of 26.8 million people who are differentlyabled. This is roughly 2.21 percent in percentage terms. Out of the total disabled person, 69% reside in rural areas whereas 31% in urban areas. [l] There are various challenges faced by the specially-abled people for Health Facilities, Access to Education, Employment Facilities and the Discrimination/ Social Exclusions top it all. Sign Language is commonly used to communicate with deaf people. Sign Language is one of the greatest adaptation for persons with speech and hearing impairments. It is also known as a visual language. There are more than 120 distinct sign languages, such as American sign language, Indian Sign Language, Italian sign language, etc

II. LITERATURE SURVEY

We examine a few similar systems that have been explored and implemented by other researchers in order to have a better understanding of their methods and strategies. Smart Glove For Deaf And Dumb Patient[3], The author's objective in this paper is to facilitate human beings by way of a glovebased communication interpreter system. Internally, The glove is attached to five flex sensors and is fastened. For each precise movement, the flex sensor generates a proportionate change in to process these hand motions. It's a combination of a microcontroller and the LA7BV/EW software that's been improved. It compares the input signal to memory-stored specified voltage values According to this, a speaker is used to provide the appropriate sound.

Sign Language Recognition[5], In 2016, proposed a unique approach to assist persons with vocal and hearing difficulties in communicating. This study's authors discuss a new method for recognizing sign language and translating speech into signs. Using skin colour segmentation, the system developed was capable of retrieving sign images from video sequences with less crowded and dynamic histories. It can tell the difference between n static and dynamic gestures and extract the appropriate feature vector. Support Vector Machines are used to categorise them. Experiments revealed satisfactory sign segmentation in a variety of backdrops, as well as fairly good accuracy in gesture and speech recognition.

Real-Time Recognition of Indian Sign Language[6], The authors have designed a system for identifying Indian sign language motions in this work. (ISL). The suggested method uses OpenCV's skin segmentation function to locate and monitor the Region of Interest (ROI). To train and predict hand gestures, fuzzy c-means clustering machine learning methods are utilised. The proposed system, according to the authors, can recognize realtime signs, making it particularly useful for hearing and speech-challenged individuals to communicate with normal people.

Deep Learning for Sign Language Recognition on Custom Processed Static Gesture Images[8], The outcomes of retraining and testing this sign language are presented in this research. Using a convolutional neural network model to analyse the gestures dataset Inception v3 was used. The model is made up of several parts. Convolution filter inputs are processed on the same input. The accuracy of validation attained was better than 90%. This is a paper describing the multiple attempts at detecting sign language images using machine learning and depth data.

III. CONVOLUTION NEURAL NETWORK

The Convolution Neural Network (CNN) is a deep learning method inspired by human neurons. A neural network is a collection of artificial neurons known as nodes in technical terms. A neuron in simple terms is a graphical representation of a numeric value. These neurons are connected using weights(numerical values). Training refers to the process where a neural network learns the pattern required for performing the task such as classification, recognition, etc. When a neural network learns, the weight between neurons changes which results in a change in the strength of the connection as well. A typical neural network is made up of various levels. The first layer is called the input layer, while the output layer is the last. In our case of recognizing the image, this last layer consists of nodes that represent a different class. We have trained the model to recognize Alphabets A to Z & Numerals 0 to 9. The likelihood of the image being mapped to the class represented by the node is given by the output neuron's value. Generally, there are 4 layers in CNN Architecture: the convolutional layer, the pooling layer, the ReLU correction layer, and the fullyconnected layer.

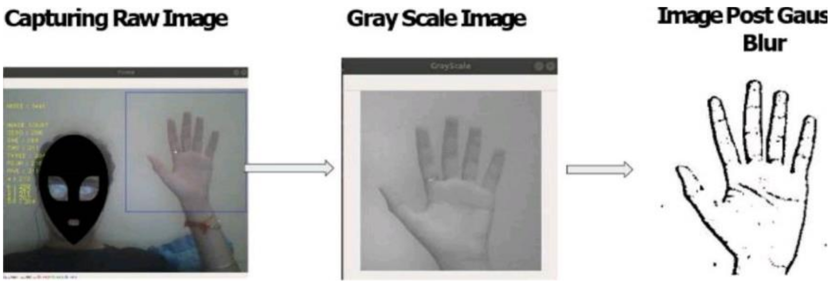
The Convolutional Layer is CNN's first layer, and it works to detect a variety of features. Images are fed into the convolutional layer, which calculates the convolution of each image with each filter. The filters match the features we're looking for in the photographs to a match. A feature map is created for each pair (picture, filter). The pooling layer is the following tier. It takes a variety offeature maps as inputs and applies the pooling method to each of them individually. In simple terms, the pooling technique aids in image size reduction while maintaining critical attributes. The output has the same number of feature maps as the input, but they are smaller. It aids in increasing efficiency. and prevents over-learning. The ReLU correction layer is responsible for replacing any negative input values with zero.lt serves as a mode of activation. The fully connected layer acts as the final layer. It returns a vector with the same size as the number of classes the image must be identified from. Mobilenet is a CNN Architecture that is faster as well as a smaller model. It makes use of a Convolutional layer called depth-wise separable convolution. Sign Language Recognition[5], In 2016, proposed a unique approach to assist persons with vocal and hearing difficulties in communicating. This study's authors discuss a new method for recognizing sign language and translating speech into signs. segmentation.

IV. SIGN LANGUAGE & ITS CONTRIBUTION

Sign Language was discovered to be a helpful way of communication since it used hand gestures, facial emotions, and mild bodily movements to transmit the message. It is extremely important to understand and interpret the sign language and frame the meaningful sentence to convey the correct message which is extremely important and challenging at the same time. The purpose of this work is to contribute to the field of sign language recognition. Humans have been trying hard to adapt to these sign languages to communicate for a long time. Hand gestures are used to express any word or alphabet or some feeling while communicating. Sign Language Recognition is a multidisciplinary subject on which research has been ongoing for the past two decades, utilising vision-based and sensorbased approaches. Although sensor-based systems provide data that is immediately usable, it is impossible to wear dedicated hardware devices all of the time. The input for vision-based hand gesture recognition could be a static or dynamic image, with the processed output being either a text description for speech impaired people or an audio response f for vision-impaired people. In recent years, we have seen the involvement of machine learning techniques with the advent of Deep Learning techniques contributing as well. A dataset is an essential component of every machine learning program. We can 't train a machine to produce accurate results without a good dataset. We created a dataset of Sign language images for our project. The photos were taken with a variety of backgrounds. After collecting all of the photos, they were cropped, converted to RGB channels, and labelled. The benefit of this is that the image size and other supplementary data are minimised allowing us to process it with the fewest resources possible.

V. METHODOLOGY

In proposals to solve the problem of gradient vanishing, an RNN structure (using forgetting units such as long short-term memory (LSTM) and gated recurrent units (GRU)) is usually proposed to enable the storage units. Such a mechanism is used to determine when to forget certain information and determine the optimal time delay. The methodology proposed in this research employs a combination of GRU and LSTM; given the known features of these methods, a combination may be expected to detect and identify sign language gestures from a video source and to generate the associated English word. In this method, the video source is provided as an input, consisting of ISL sign-language gesture sequences. The primary purpose of this system is to identify the words from signing gestures used in real life. Therefore, the foremost task is to divide the video file containing the sequence of ISL gestures for various words into separate subsection videos containing different words. This is performed by identifying the start and end of each different gesture. After dividing up the videos, the next step is to divide the resulting subsection videos into frames. The frames are generated by applying sampling techniques to the video sequence. In order to extract the features from the frames, InceptionResNetV2 was used in the proposed architecture [20—22]. InceptionResNetV2 performs better when compared to its earlier variant. The video frames are resized and passed to the inceptionResNetV2 model, using mobile net pre-trained weights to extract the features from the gestures. The output of the inceptionResNetV2 model is an array consisting of the feature vector. These features are passed to a recurrent neural network to predict the correct word.



Gaussian

VI. CONCLUSION

A technique for recognizing Indian Sign Language is presented in this work. The Tensorflow Mobilenet V2 Model was used to recognize static indicators successfully. The Model can further be improved by adding more numbers of signs and increasing the dynamicity of the Images. A functional real time vision based American sign language recognition for deaf and dumb people have been developed for ASL alphabets. We can achieve final accuracy of 98.74 percent on our data set. We would be able to improve our prediction after implementing two layers of algorithms in which we verify and predict symbols which are more similar to each other. This way we are able to detect almost all the symbols provided that they are shown properly, there is no noise in the background and lighting is adequate. ; this approach could be utilized for interpreting continuous sign language that leads to syntax generation, especially in the context of ISL. The use of vision transformers can lead to more accurate results than those of feedbackbased learning models. Author Contributions: D.K.: Data Curation, Investigation, Writing—original draft, C.B.: Methodology, Project administration, Writing—review and editing, K.S.: Data curation, Investigation, Software, K.P.: Data curation, Investigation, Software, A.-B.G.-G.: Methodology, Writing —review and editing, J.M.C.: Methodology, Writing—review and editing. All authors have read and agreed to the published version of the manuscript. Funding: This work has been supported by the Institute for Business Competitiveness of Castilla y Leon and the European Regional Development Fund under grant CCTT3/20/SA/0002 (AIR-SCity project). Data Availability Statement: The dataset can be found here:

https://github.com/DeepKothadiya/ Custom ISLDataset/tree/main (the authors ' permission is required when accessing the dataset) accessed on 25 March 2022. Acknowledgments: Our past UG students played essential role for dataset generation. We are thankful to our students Riya Patel, Yashvi Raythatha and Siddharth Patel for their contribution and active involvement in dataset generation. Conflicts of Interest: The authors declare no conflict of interest.

VII. REFERENCES

[1] Ministry of Statistics & Programme Implementation. Available online: https://pib.gov.in/PressReleasePage.aspx?P RID=1593253 (accessed on 5 January 2022)

[2] Manware, A.; Raj, R.; Kumar, A.; Pawar, T. Smart Gloves as a Communication Tool for the Speech Impaired and Hearing Impaired. Int. J. Emerg. Technol. Innov. Res. 2017, 4, 78-82

[3] Wadhawan, A.; Kumar, P. Sign language recognition systems: A decade systematic literature review. Arch. Comput. Methods Eng. 2021, 28, 785-813. [CrossRef]

[4] Papastratis, 1.; Chatzikonstantinou, C.; Konstantinidis, D.; Dimitropoulos, K.; Daras, P. Artificial Intelligence Technologies for Sign Language. Sensors 2021, 21, 5843. [CrossRef] [PubMed]

[5] Nandy, A.; Prasad, J.; Mondal, S.; Chakraborty, P.; Nandi, G. Recognition of Isolated Indian Sign Language Gesture in Real Time. Commun. Comput. Inf. Sci. 2010, 70, 102-107

[6] Mekala, Gao, Y; Fan, Davari, A. Real-time sign language recognition based on neural network architecture. In Proceedings of the IEEE 43rd

[7] Southeastern Symposium on System Theory, Aubum, AL, USA, 14-16 March 2011.

[8] Chen, J.K. Sign Language Recognition with Unsupervised Feature Learning; CS229 Project Final Report; Stanford University: Stanford, CA, USA, 2011.

[9] Sharma, M.; Pal, R.; Sahoo, A. Indian sign language recognition using neural networks and KNN classifiers. J. Eng. Appl. Sci. 2014,