International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 10 Issue 5 ǁ 2022 ǁ PP. 67-71

"Real Time Translation of Sign Language to Text"

[1]Swathi Sri Parthiban
[1]PG Student, Department of CS and IT,Jain (Deemed-to-be University)

ABSTRACT
Making a work area application that utilizes a computer webcam to catch an individual appearance signal for Indian Sign Language (ISL), and make an interpretation of it into relating text and discourse progressively. The deciphered communication through signing motion will be addressed in the text. A Fingerspelling gesture-based communication interpreter is being proposed. To empower the discovery of motions, the Convolutional Neural Network (CNN) is applied. CNN is proficient in handling computer vision issues and fit for distinguishing the ideal highlights with a serious level of exactness upon adequate preparation. The proposed framework is to change over the hand token of communication via gestures to message Using AI strategies. A framework is invested to catch a genuine energy interpretation of Indian gesture-based communication utilizing a single hand.

KEYWORDS:
Indian Sign Language, Hand Gesture Recognition, Convolution Neural Network, K- means algorithm, Open CV.

Date of Submission: 12-05-2022 Date of acceptance: 26-05-2022

I. INTRODUCTION

Physically impaired individuals use hand signs and motions to convey. Common people face trouble in understanding their language. Henceforth there is a need for a framework that understands the various signs, motions and passes the data on to the common individuals. It overcomes any barrier between actually tested individuals and ordinary individuals.

Deep learning is an Artificial Intelligence(AI) work that copies the working of a human mind in handling information and making approach-designs for use in navigation. Deep Learning is a subset of AI in computerized reasoning that has network effectiveness of gaining solo from information that is changeable or unlabelled. Deep Learning anyway known as deep network learning is important for a boundary group of AI strategies worked of counterfeit brain networks with portrayal learning. Deep Learning then again is a kind of Machine learning, which is arouse by the construction of a human mind. The primary benefit of profound realizing all-over AI is the unnecessariness of the supposed component extraction. The consequence of Feature Extraction is a portrayal of the given crude information that can be used by these exemplary AI calculations to perform assignments. For instance, the arranging of the information into a few classifications or classes. Include extraction is typically very confounded and includes definite information on the issue area. This pre-processing layer should be coordinated, tried, and refined with more than a few emphases for ideal outcomes.

INDIAN SIGN LANGUAGE DETECTION: Human beings can express their sign languages in various ways, such as facial expression, bodily expressions, or language. Sign Language Recognition (SLR) intricacies are viewed as normal in many individuals and this will be a test, as understanding gesture-based communication further develops individuals to act in a socially notable way. Communication through signing can be recognized from an assortment of references of descriptor, including looks, manner of speaking, and body act. These measures incorporate slow practices and thin interests, as well as liabilities in friendly conduct and postponements in verbal and non-verbal correspondence. Moreover, kids with hearing lack in their expressive language and weaknesses in correspondence. This could prompt an entire absence of language advancement and prompt delay in communication in the language. The Voice problem and hearing misfortune youngsters disturbed in knowing about others language and social collaborations. These attributes incorporate lacking eye-to-eye connection, look, non-verbal parts of correspondence, and body signal. This additionally prompts disappointment in sharing interests with others. Social debilitations might reflect as an absence of social correspondence and knowledge into their own and others' communication through signing. The limited, redundant, and generalized practices are the third kind of quality to analyse the Voice issue and hearing misfortune Disorders. The presentation measure can be determined by utilizing quantitative investigation. This finding signalizes that youngsters have slow simplicity in perceiving the accompanying gesture-based communication, for example, irate, dread, blissful, miserable, impartial, and ordinary correspondences.

II. LITERATURE SURVEY

Tanuj Bohra et al. [1] proposed a real-time two-way sign language communication system built using image processing, deep learning and computer vision.hand detection, skin segmentation, contour detection is performed on images which gives accuracy of 99%. Joyeeta Singha and Karen Das [2] proposed a system for Indian sign language recognition from a live video. The system comprises of three stages like pre-processing, extraction and classification and achieved an accuracy of 96.25%. Muthu Mariappan H. et al. [3] designed a real time sign language recognition system as a portable unit using contour detection and fuzzy c-means algorithm. contour is used to detecting face, hand and fuzzy c-mean algorithm is used to partition input data to specifies clusters and achieved an accuracy of 75%Salma Hayani et al. [4] proposed an Arab sign language recognition system based on CNN, inspired from LeNet-5 and the results are compared with the machine learning algorithm like KNN and SVM to show the performance of the system and 90%accuracuy is gained. Kshitij Bantupalli and Ying Xie [5] worked on American sign language recognition system which works on video sequences based on CNN, LSTM and RNN. A CNN model named Inception was used to extract spatial features from frames, LSTM for longer time dependencies and RNN to extract temporal features and maximum accuracy of 96% is obtained. Mahesh Kumar [6] proposed a system which can recognize 26 hand gestures of Indian sign language based on Linear Discriminant Analysis (LDA). Pre-processing steps such as skin segmentation and morphological operations are applied on the dataset. Skin segmentation was carried out using Otsu algorithm. Linear discriminant analysis is used for feature extraction. Suharto et al. [7] tried to implement a sign language recognition system with I3D inception model through transfer learning method. Public dataset LSA64 was used for 10 vocabularies with 500 videos. The model has good training accuracy but very low validation accuracy. Oscar Kollar et al. [8] introduced a hybrid CNN-HMM for sign language recognition. They conducted experiments on three datasets namely RWTH-PHOENIX-Weather 2012, RWTH-PHOENIX-Weather Multisigner 2014 and SIGNUM single signer. Mengyi Xie and Xin Ma [9] proposed an end-to-end system using residual neural network to implement recognition of American sign language. The images are converted to CSV file format and after applying one-hot encoding are given as input to ResNet50 network for training. Model gives an accuracy of 96.02%. G. Anantha Rao et al. [10] proposes an Indian sign language gesture recognition using convolutional neural network. This system works on videos captured from a mobile 's front camera. Average recognition rate of this CNN model is 92.88%. Aditya Das et al. [11] trained a convolutional neural network using Inception v3 model for American sign language. Data augmentation is applied on the images before training 12 them to avoid overfitting. This model gives more than 90% accuracy on Sreehari sreejith dataset for 24 class labels with 100 images per class.

III. METHODOLOGY

The proposed work uses CNNs with variable depths to evaluate the performance ofthe model. The following network architectureis considered in the model: [Conv-(SBN)- ReLU-(Dropout)- (Max-pool)] M - [Affine-(BN)- ReLU-(Dropout)] N - Affine – Softmax. The initial segment of the network refers to M convolutional layers that can have Spatial Batch Normalization (SBN), dropout, and max-pooling in addition to the convolution layer and ReLU nonlinearity, which generally exists in these layers. After the M convolution layers, the network is directed to N fully connected layers that generally have Affine activity and ReLU nonlinearity, and can incorporate batch Normalization (BN) and dropout. At last, the network is trailed by the relative layer that registers the scores and softmax loss function. The proposed model gives the user the opportunity to choose the number of convolutional and fully connected layers, as well as the presence of batch normalization, dropout, and max-pooling layers. In addition, with dropout and batch normalization techniques, the L2 regularization is carried out. In addition, the amount of channels, strolls, and zero-padding can be specified by the user, and if they are not given, the default values are considered.

Model Construction

A convolutional neural network is a multilayer perceptron which is used for pattern classification and it is encouraged by the natural visual perception mechanism of living thing. Such networks usually consist of an input, an output and multiple hidden layers. Hidden layers typically consist of convolution, pooling, fully connected and normalization layers. Convolutional neural networks are used for various applications, including image classification, object detection, object tracking, text detection and recognition, speech and natural language processing. The emotion recognition model is based on a deep learning approach, which uses convolution neural networks(CNNs) which is one of the most effective deep learning structures. The CNN model is being trained and prediction of image classification is conducted by the input. This prediction is compared to the images label and the loss of the estimation is regulated. The data collection interface was created by first fine-tuning a CNN model that was pre-trained on an existing dataset of face expression images which is been collected from the web. The users face images which is been collected from the device are then analysed by the CNN model, labelled and added to the new dataset. By this approach, an approximately a dataset of 20,000 images is collected, which were used to train a fine-turned CNN model, which forms the basis for the Emotional Training Platform.

Framework:

To achieve successful Sign Language computing performance in real-time, a framework is implemented. The represented gesture expression appears on the screen. By integrating the work into the recognition of sign language in real-time from a web camera, the platform is capable of tracking the gestures of the user in real-time and assessing whether the user makes these gestures with precision. Training in sign language expression is replicating to be necessary to improve recognition skills. This kind of training allows us to enhance people's skills with recognition of hand gestures in Indian Sign Language.

IV. RESULT AND DISCUSSION

The current work which is been proposed gives the expected outcome when recognised sign language are given. The result will be different for each motion which is been captured by the webcam and furthermore relies upon the individual representing the sign language gesture. This work reads numerous signals consistently and gives the particular result which is been captured by the web camera.

The proposed system's strategy is that the webcam identifies the sign gesturewhich is been shown in the camera and detect the number of noise and imperfections which is present in the gesture. The exact sign is beingidentified by the outline of the hand and the noise is being avoided to get the accuracy. The result is being represented in the form of text or number based on the image captured.

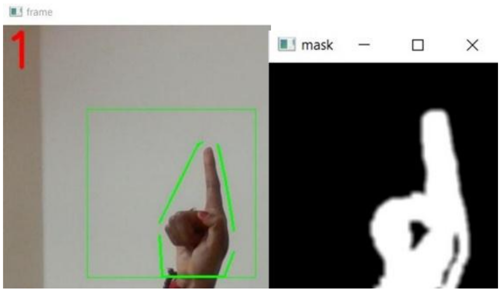

Fig 1: SignlanguageforONEwithmask

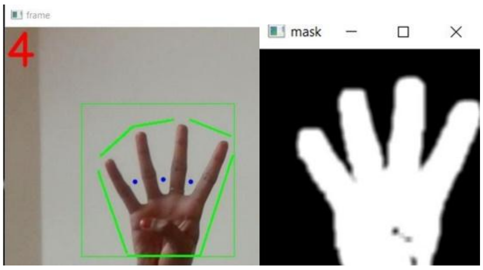

Fig 2: SignlanguageforFOURwithmask

Fig 3: SignlanguageforOKwithmask

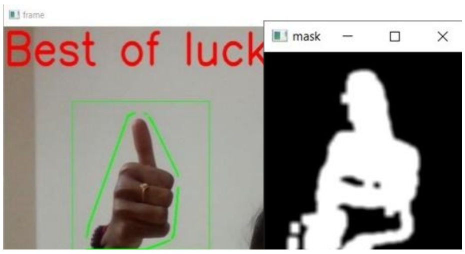

Fig 4: SignlanguageforBEST OF LUCKwithmask

Fig 1 to Fig 4 illustrate the Sign Language for hand gestures where the frame represents the value of the image captured by the web camera and the mask represent the image with the noise and blurring the background by representing it in white and black for better accuracy.

V. CONCLUSION

The proposed work is a small step towards helping a physically challenged people and lot more can be done to make the product more sophisticated, user friendly and efficient. This proposed work aims to predict sign language recognition using machine vision with the help of deep learning. The performance of this tool is on part with human for distinguishing the sign language gesture with real time image. It recognizes the hand gestures with probably Indian Sign Language capturing the real-time image and then training the data of hand gestures and converting it to text. The project aim is to even predict the recognition of hand gestures in online mode i.e. Google meet and Zoom meet etc. the system is successfully designed and implemented a sign language interpretation system for Indian Sign language.

The proposed system can be enhanced by the means of bi directional responses, which can further be improved to convert the speech to sign language and sing language to speech. This translator device helps the physically challenged people to socially connect with the people in real time. This real time sign language detection system can be further improved to recognise the motion of the hand like rotation, clapping etc. This system can also be extended by sign recognition to text and speech and vice versa simultaneously with more efficient algorithm. The system can also be enhanced to display multi-languages and converting them to speech.

REFERENCES

[1]. Bohra, T., Sompura, S., Parekh, K., & Raut, P. (2019, November). Real-Time Two Way Communication System for Speech and Hearing Impaired Using Computer Vision and Deep Learning. In 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 734-739). IEEE.

[2]. Singha, J., & Das, K. (2013). Recognition of Indian sign language in live video. arXiv preprint arXiv:1306.1301.

[3]. Mariappan, H. M., & Gomathi, V. (2019, February). Real-time recognition of Indian sign language. In 2019 International Conference on Computational Intelligence in Data Science (ICCIDS) (pp. 1-6). IEEE.

[4]. Bchir, O. HAND SEGMENTATION FOR ARABIC SIGN LANGUAGE ALPHABET RECOGNITION.

[5]. Bantupalli, K., & Xie, Y. (2018, December). American sign language recognition using deep learning and computer vision. In 2018 IEEE International Conference on Big Data (Big Data) (pp. 4896-4899). IEEE.

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

[7]. Gunawan, H., Thiracitta, N., & Nugroho, A. (2018, September). Sign language recognition using modified convolutional neural network model. In 2018 Indonesian Association for Pattern Recognition International Conference (INAPR) (pp. 1-5). IEEE.

[8]. Koller, O., Zargaran, O., Ney, H., & Bowden, R. (2016). Deep sign: Hybrid CNN-HMM for continuous sign language recognition. In Proceedings of the British Machine Vision Conference 2016.

[9]. MADHUBALA, D., RAO, N. K., ROOPA, D. Y. M., KALYANI, D., PRANAY, M., & RAJU, C. K. A TOOL TO CONVERT AUDIO/TEXT TO SIGN LANGUAGE USING PYTHON LIBRARIES. Turkish Journal of Physiotherapy and Rehabilitation, 32, 2.

[10]. Alnaim, N., Abbod, M., & Albar, A. (2019, October). Hand Gesture Recognition Using Convolutional Neural Network for People Who Have Experienced A Stroke. In 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) (pp. 1-6). IEEE

[11]. Das, A., Gawde, S., Suratwala, K., & Kalbande, D. (2018, January). Sign language recognition using deep learning on custom processed static gesture images. In 2018 International Conference on Smart City and Emerging Technology (ICSCET) (pp. 1-6). IEEE

[12]. Elizabth Cody, "Deaftravel", http://www.deaftravel.co.uk/signs.php?id=27

[13]. Richmond Alake, "Understanding and Implementing LeNet-5 CNN Architecture", June 2020, https://towardsdatascience.com/understanding-and-implementing-lenet-5-cnn-architecture-deep-learning-a2d531ebc342

[14]. Priyankur Sarkar, "What is LDA: Linear Discriminant Analysis for Machine Learning", September 2019, https://www.knowledgehut.com/blog/data-science/linear-discriminant-analysis-for-machine-learning

[15]. Ronchetti, F., Quiroga, F., Estrebou, C., Lanzarini, L., & Rosete, A. (2016). Lsa64: A dataset of argentinian sign language. In XX II Congreso Argentino de Ciencias de la Computación (CACIC).

[16]. Madeline Schiappa, "Understanding the Backbone of Video Classification: The I3D Architecture", June 2020, https://towardsdatascience.com/understanding-the-backbone-of-video-classification-the-i3d-architecture-d4011391692

[17]. Priya Dwivedi, "Understanding and Coding a ResNet in Keras", January 2019, https://towardsdatascience.com/understanding-and-coding-a-resnet-in-keras-446d7ff84d33

[18]. Sik-Ho Tsang, "Review: Inception-v3-1 st Runner Up (Image Classification) in ILSVRC 2015", September 2018, https://shtsang.medium.com/review-inception-v3-1st-runner-up-image-classification-in-ilsvrc-2015-17915421f77c.

[19]. Forster, J., Schmidt, C., Hoyoux, T., Koller, O., Zelle, U., Piater, J., & Ney, H. (2012, May). Rwth-phoenix-weather: A large vocabulary sign language recognition and translation corpus. In Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12) (pp. 3785-3789).

[20]. Koller, O., Forster, J., & Ney, H. (2015). Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers. Computer Vision and Image Understanding, 141, 108-125.