International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 05, May-2022

Real Time Static and Dynamic Hand Gesture Recognition using CNN

Amrutha D
Bhumika M
Shivani Hosangadi, Shravya
Students, BMS Institute of Technology and Management

Manoj. H. M
Assistant Professor, Dept. Of CSE,
BMS Institute of Technology and Management,
Bengaluru, Karnataka, India

***Abstract*:- Sign language is a communication method for hearing disable people. People with hearing disabilities face a lot of problems communicating with other hearing people without a translator. In this project, we have proposed a marker-free, visual American Sign Language recognition system using image processing, computer vision and neural network methodologies. This approach will convert hand gestures into a text. A sign language translator is an important way for communication with the deaf people and the general public. So here we are doing development and implementation of an American Sign Language (ASL) hand gestures translator based on a Convolutional neural network. We produced a CNN(Convolutional Neural Network) model which trains and categorizes alphabet letters a-z and translates into its respected text in a majority of cases. Various Neural network algorithms are applied on the datasets, including RNN (recurrent Neural Network). An attempt is made to increase the accuracy of the CNN model by pre-training it on the dataset.**

Keywords : Hand Gesture Recognition, Sign Language, CNN, RNN, Spatial and Temporal Sign Language

1. INTRODUCTION

A very important requirement for social survival is communication. More than 5 to 10% of today's population are suffering from hearing or speaking problems. So to overcome this problem and to facilitate easy communication for such individuals sign languages are developed. Sign language is mainly used by deaf and dumb people to communicate with one another. But the main problem is that the normal people find it difficult to understand. Basically American Sign Language (ASL) is different from British Sign language (BSL) and Indian Sign Language((ISL) in a variety of ways. To communicate in ISL, BSL both hands are used whereas in ASL only problem is recognition that there may be some problem in recognition that it may fail to recognize some similar gestures. Our project mainly focuses on recognizing the gestures which are alphabet based and some English word based.

In this project we mainly focused on American Sign Language Recognition. We mainly used CNN which is the basic method in deep learning. So as to overcome the recognition problem we use prestored datasets. We use laptop or smartphone cameras to capture the images created in the datasets. The proposed methodology is based on Static and Dynamic hand gestures. When a new dataset is created in the field of deep learning it may be a new contribution to that field just because each and every dataset has its own special features so as to improve the existing models.

There are two types of gestures in American Sign Language Recognition, namely Static Recognition Gestures also called as Spatial Hand Gestures and Dynamic Recognition Gestures also called as Temporal Hand Gestures. Static Recognition Gestures is one of the most significant of sign language recognition as it is used in a number of situations such as addresses, brands, names, etc. There may be some visual similarities in different signs, hence this static Recognition Gesture is difficult compared to Dynamic Recognition Gesture. Apart from this, there are a number of variations which depend on the viewpoint of the camera. The main advantages of using deep learning with CNNs is to overcome this problem and to achieve real time. Also it helps in achieving an accurate sign finger spelling recognition model.

The National Institute of Deafness and Other Communications Disorders (NIDCD) has proved that the 200-year-old American Sign Language is a complex language but is considered as the primary language for many deaf and dumb. Hence to build a system which can be used to recognize sign language will help the deaf and dumb for easy communication modern-day technologies.

2. RELATED WORKS

In recent years, Many researchers designed different methods of gesture recognition in different fields. Recognition maybe vision based, glove based, Artificial Neural Network based, or based on soft computing approaches like Artificial Neural Network, Fuzzy Logic, PCA etc. Literature review of our proposed method shows that there have been many explorations done to get the sign language recognition in images using various methods and algorithms.

The research was done by Usha which made use of the YCbCr skin model to detect and fragment the skin region of the hand gestures. Using Principal that are based on Region Detector, the picture features are extracted and classified with using Multi class SVM and non-linear KNN. A CNN is used for hallmark-extraction and sign language recognition consisting of Long Short - Term Memory (LSTM) coding network are built for the language image capturing.
On the problem of sign language image recognition in practical problems , the paper get the hand locating network,CNN features extraction network encoding and decoding to construct various algorithms for extraction. This paper has done a recognition of 99% in vocabulary dataset.

A dataset of Indian Sign Language static alphabet signs were used for training. The results obtained were 94.4%for static.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 05, May-2022

A budgetary cost approach has been used for image processing and extraction. The capturing of images will be done with a green and gray background so that during processing, the green color can be easily debited from the color space and the image gets metamorphosed to black and white.

The methods that have been proposed in the study have mapped the signs using the centroid method. It can classify input gestures with a database having various hand sizes and various positions. The prototype has accurately recognised 93% of the sign gestures.

The paper by M. Geethanjali and U. M. Shalini, makes use of 50 specimens of every letter recognition of American Sign Language characters using C-Spine approximations. The boundary which is obtained is further converted to an B-spline by making use of the Maximum Curvature Points(MCPs).

The B-spline curve has to undergo a series of smoothening processes as features can be extracted. Support vector machine(SVM) is used to classify the images and the accuracy is 92.01%. Pigouge used CLAP15 as a dataset. It consists of 26 American sign gestures. After image-preprocessing the pictures , he used Convolutional Neural network(CNN) model having 7 layers for training. It is to be noted that the model is not a 3D CNN and all the kernels are in 3D. He has used Rectified linear Units (ReLU) as activation methodology. Feature extraction done using CNN while classification uses RNN or fully connected layers. His work has achieved an accuracy of 93.70% with an error rate of 7.30%.

A similar work has been done by M Huang. He created his own dataset by using Kinect and got a total of 28 vocabulary words which are used in everyday lives. He then applied a 3D RNN in which most of the kernels are also in 3D. The input and output of his model consisted of 5 important channels which are color-b,color-r, color-g, depth and body skeleton. He got an accuracy of 95.2%.

Another research paper on Action and method recognition topic by the author N.Carriera shares some of the similarities to sign gesture recognition.He used a transfer learning method and action for his research As his trained dataset, he used ImageNet. After training the models using another three datasets namely UCF-102 and HMDB-53 , he then came up with the RGB model, flow model, pre-trained Kinetic and pre-trained ImageNet.The accuracy he got on UCF-102 dataset is 97.5%.

3. METHODOLOGY

In the proposed technique, first we are able to extract spatial features for individual frames by the usage of inceptionv3 model (CNN) and temporal features by the usage of RNN. Each video (a series of frames) is then represented by a sequence of predictions made via way of means of CNN for each of the individual frames. This collection of predictions was given as enter/input to the RNN.

- First, we are able to extract the frames from the more than one video sequences of every gesture.
- After the first step, noise from the frames i.e. background, frame elements apart from hand gestures are eliminated/removed to extract more applicable features from the frame.
- Frames of the train data are given to the CNN model for training on the spatial features. We have used inception model for this purpose which is a deep neural net.
- Store the train and test frame predictions. We'll use the model obtained in the above step for the prediction of frames.
- The predictions of the trained data are now given to the RNN model for training on the temporal features. We have used LSTM (Long short - term memory) model for this purpose.

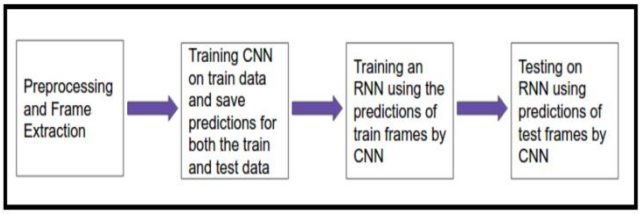

Fig 3.1 :Methodology

3.1 FRAME EXTRACTION AND BACKGROUND REMOVAL

Each video gesture video is broken down into a sequence of frames. Frames are then processed to remove all the noise from the image that is everything except hands. The final image consists of grey scale image of hands to avoid colour specific learning of the model.

Fig. 3.1.1a : One of the extracted frame

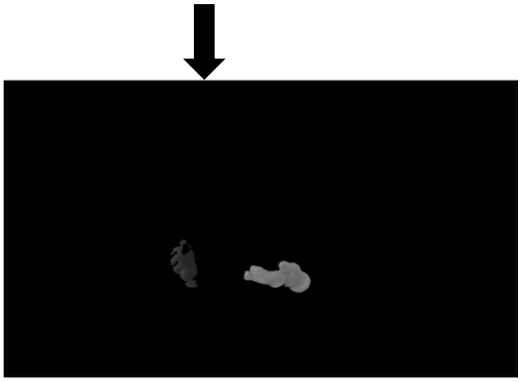

Fig 3.1.1b: Frame after extracting hands (Background Removal)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 05, May-2022

3.2 TRAIN CNN(SPATIAL FEATURES) AND PREDICTION

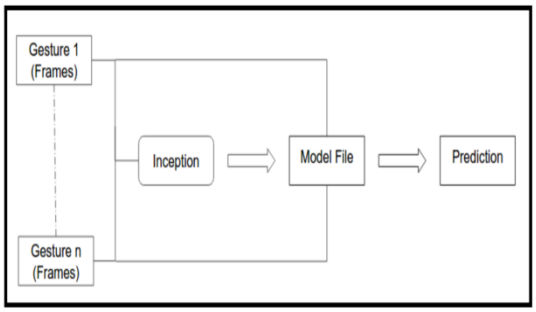

Fig 3.2.1 : CNN Training

The set of frames from the input video of gestures will be extracted and sequence of predictions for each frame will be displayed by CNN after training it.

3.3 TRAINING RNN (TEMPORAL FEATURES)

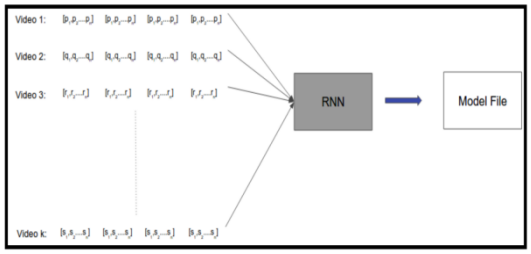

Fig 3.3.1 : RNN Training

The sequence of predictions by CNN for temporal features will be given to RNN for training. These features will be trained using RNN. RNN produces Model file , which will be used for prediction of temporal images.

4. EXPERIMENTAL RESULTS

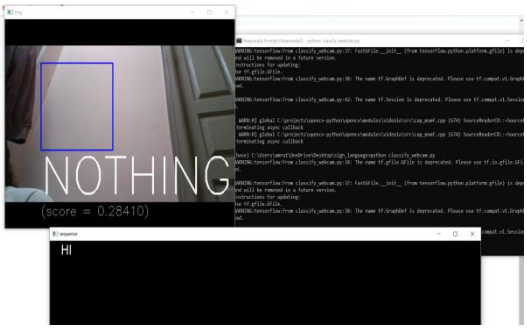

Fig 4.1 : Final User Interface with output

5. CONCLUSION

Vision based hand gesture recognition techniques have many proven advantages compared with traditional devices. However, hand gesture recognition and its conversion to text or speech is a difficult problem and the current work is only a small contribution towards achieving the results needed in the field of sign language gesture recognition. This report presented a vision based system able to interpret isolated hand gestures from the American Sign Language(ASL).

Videos are difficult to classify because they contain both the temporal as well as the spatial features. We have used two different models to classify on the spatial and temporal features. CNN was used to classify
on the spatial features whereas RNN was used to classify on the temporal features. We obtained an accuracy of 95.217%. This shows that CNN along with RNN can be successfully used to learn spatial and temporal features and classify Sign Language Gestures.

The current process uses two different models, training inception (CNN) followed by training RNN. For future work one can focus on combining the two models into a single model.

6. REFERENCES

[1] Mehreen Hurroo , Mohammad Elham , “Sign language Recognition System using Convolutional Neural Network and Computer Vision” , Volume 09, Issue 12 (December 2020), IJERT.
[2] Christian Zimmermann, Thomas Brox , “Learning to Estimate 3D Hand Pose from Single RGB Images”, ICCV 2019.
[3] Huang, J., Zhou, W., & Li, H , “Sign Language Recognition using 3D convolutional neural networks”, IEEE International Conference on Multimedia and Expo, 2015.
[4] Pratibha Pandey, Vinay Jain, “Hand Gesture Recognition for Sign Language Recognition”, A Review, International Journal of Science, Engineering and Technology Research (IJSETR), Volume 4, Issue 3, March 2015.
[5] Fares Ben Slimane, Mohamed Bouguessa Hongdong Li, “Self-Attention for Sign Language Recognition”, Context Matters, 12 Jan 2021.
[6] Sunitha K. A, Anitha Saraswathi.P, Aarthi.M, Jayapriya. K, Lingam Sunny, “Deaf Mute Communication Interpreter” , A Review, International Journal of Applied Engineering Research, Volume 11, pp 290-296, 2016.
[7] Dongxu Li, Chenchen Xu, Xin Yu, Kaihao Zhang, Ben Swift, Hanna Suominen, “Hierarchical Feature Learning via Temporal Semantic Pyramid for Sign Language Translation” , TSPNet, NeurIPS 2020
[8] Mandeep Kaur Ahuja, Amardeep Singh, “Hand Gesture Recognition Using PCA”, International Journal of Computer Science Engineering and Technology (IJCSET), Volume 5, Issue 7, pp. 267-27, July 2015.
[9] G Ananth Rao and PVV Kishore. “Sign language recognition system simulated for video captured with smart phone front camera”. International Journal of Electrical and Computer Engineering,6(5):2176, 2016.
[10] G Ananth Rao and PVV Kishore. “Selfie video based continuous Indian sign language recognition system”. Ain Shams Engineering Journal,2017.
[11] Zhengzhe Liu, Fuyang Hyung, Gladys Wai Lan Tang, Felix Yim Binh Sze, Jing Qin, Xiaogang Wang and Qiang Xu. “Real-time sign language recognition with guided deep convolutional neural networks. In Proceedings of the 2016 Symposium on Spatial User Interaction, pages 187– 187. ACM, 2016.