IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
Research paper **Volume 11, Sp.Iss 5 , 2022**

INFLATED 3D CONVNET FOR DETECTION OF SIGN LANGUAGE

Sridevi Sakhamuri[1] , D Chandana[2], M Tushara[3], A Ramya Sri[4],

[1] Assistant Professor, Department of IoT, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram,Guntur Dist, Andhra Pradesh, 522302, India

[2,3,4] Student, Department of ECM, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram,Guntur Dist, Andhra Pradesh, 522302, India

DOI : 10.48047/IJFANS/11/Sp.Iss5/057

Abstract: Every human has their own kind of disabilities, we all try to live and overcome them in our life. We educate ourselves to overcome them, we invent technology to achieve our goals. Sign Language - A way of communication for deaf and mute people through hand gestures and actions. Sign Language helps people who can't speak sign language to communicate with the people who can speak sign language,this deep learning paper aims to help build a communication bridge for this reason. We used Amazon Rekognition service which uses Deep CNN algorithm for the detection of static images of the signs. As most of the signs are for words they are in the form of videos. We used the I3D algorithm for the classification of videos of the signs of words. The PyTorch framework provides support for CuDNN (NVIDIA CUDA Deep Neural Network) which provides fast GPU implementations for the deep neural networks. The experimental results has shown that the models used has displayed good results in detecting the words.

KEYWORDS : CNN, ASL, I3D, ConvNet, CUDA, GPU, WLASL

1. INTRODUCTION

Sign language is the way groups of deaf people started communicating with each other from the 17th Century. It allows them to express themselves just like we do the same through speech. It is composed of a system of gestures, hand signs, finger spellings etc. Similar to how we communicate through different languages , Sign language is also different in different parts of the world. For example, Americans use only one hand for communicating whereas Indians use both their hands for sign language. It has also

ISSN PRINT 2319 1775 Online 2320 7876
Research paper **Volume 11, Sp.Iss 5 , 2022**

evolved with time through various interactions between deaf people[1]. There are nearly 138 to 300 different types of sign language used around the globe today. Still the main problem lies in communicating with normal people who don't learn or understand sign language which results in a communication gap between deaf people and others [2]. Detection and Recognition systemswork efficiently for people to understand the sign language which deaf people are trying to convey without actually learning the sign language. It can be useful in case of an emergency too[3].

Sign Language is defined as a system of communication using visual gestures and signs, as used by deaf people [4].Sign language detection system is basically a system which can detect a particular letter/word/sentence through an image/video input [5]. It can aid in the communication gap between a person who knows sign language and a normal person who doesn't know sign language.

Sign language isn't just for people who are born deaf [6]. We all know that with old age , there are a number of problems that come along. One of them could be loss of hearing in a period of time.On the other hand it is also a language for communication so it can be used in situations where you can't communicate orally. For example , when you are underwater, talking through a glass [7], when you are really far from the other person or when you are in a loud concert and when your mouth is full. It is also a more expressive language when compared to speech language where tone expresses your emotion.

Detection and Recognition systems are the present areas of research and development, a tremendous growth in the development of these systems are being witnessed by humans during the past decades. These systems have been very helpful in the past and are being very beneficial in the present society, they are convenient, user-friendly for the people to use [8]. Although there are many detection and recognition systems for various purposes, there are some areas where these systems have not yet delivered productive results. Sign Language Detection is one of those areas, in which some solutions are provided but have not achieved the desired results, due to which we want to provide a better solution.

The intent of this paper is to develop a Sign Language Detection System which detects the signs of words in the videos. The reasons for importance of developing is

ISSN PRINT 2319 1775 Online 2320 7876
Research paper Volume 11, Sp.Iss 5 , 2022

states previously. Most of the solutions existing currently only detect the signs in a character level from images of those signs, but in the word level the signs involved in moving the hands, each word has a different sign which is nearly 2 to 4 seconds, so the data will be videos of those signs. There are only a few algorithms which can be used for video classification and detection. Section 2 briefs about the previous Related Work, Section 3 gives the description of the Methodology used and Section 4 is the Experiments and Results, respectively.

2. RELATED WORK

As the technologies are drastically increasing day by day to provide solutions to a variety of problems in the world, there are various studies on sign language detection and multiple implementations for identifying sign language utilizing various machine learning and deep learning techniques. To understand the process of sign language detection systems, many papers have been reviewed in the past and are being reviewed currently.

Tamura et al. (1988) built a vision-based sign language recognition system by using a simple color thresholding to recognise ten isolated signs from Japanese sign language (JSL). The feature extraction was done based on the input image sequence, by tracking the hand location and then the features are extracted from the hand region. The system presented in this paper can only classify some of the movements due to the difficulty of recognising sign language words with more unrestricted hand movement[9]. Gupta et al. (2016) proposed a model that recognises static images of signed alphabets of Indian Sign Language (ISL). They used two features namely HOG (Histogram Oriented Gradient) and SIFT (Scale Invariant Feature Transform), these 2 features are extracted from the set of training images and they are combined into a single matrix. This technique is only restricted for static images and cannot be used for dynamic images or videos[10]. Rao et al. (2018) implemented a stochastic pooling technique where it unites the advantages of both max and mean pooling techniques. They used a deep CNN architecture and created 200 ISL signs with 5 signers for 2 seconds each at 30 fps (frame per second) [11].Kumar et al. (2021) used convolutional neural networks (CNNs) to recognise the American Sign Language (ASL) Alphabets. There are 3 phases; In the first phase pre-processing of MNIST dataset is done, In the second phase important features of

ISSN PRINT 2319 1775 Online 2320 7876
Research paper **© 2012 IJFANS. All Rights Reserved**, **Volume 11, Sp.Iss 5 , 2022**

pre-processed images of hand gestures are computed and in the final phase by using the properties computes in the initial phases, the accuracy of the model is detected [12].

3. METHODOLOGY

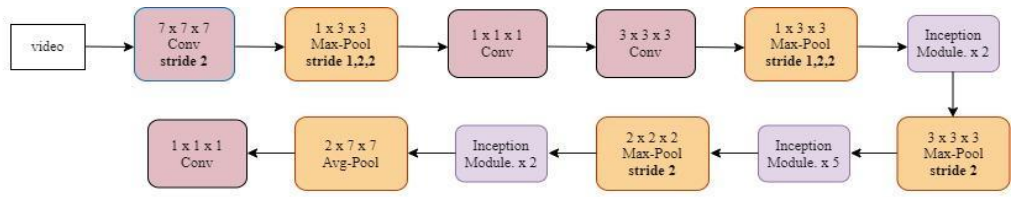
This section of the paper contains the methods used to achieve sign language detection. The same has been detailed in the upcoming subsections.

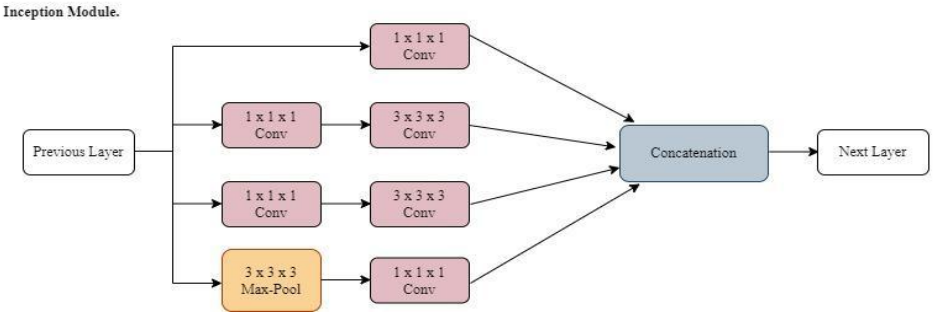
3.1. AMAZON REKOGNITION

Amazon Rekognition is a computer vision cloud service developed by Amazon Web Services. It uses deep learning technology which allows users to easily integrate image and video analysis. It uses an easy-to-use API for analysingfiles that are stored in Amazon S3 Bucket. Amazon Rekognition Custom Labels is an automated ML feature which is used to train custom computer vision models specific to the business needs quickly, just by bringing labeled images.

3.2. I3D

Inflated 3-Dimensional ConvNet (I3D) [13] was introduced for action detection in videos. This model takes a video as an input, a video here is considered as a multiple sequential 2-dimensional frames, which is a 3D input with time as a third dimension.





ISSN PRINT 2319 1775 Online 2320 7876
Research paper Volume 11, Sp.Iss 5 , 2022

Fig. 1. I3D Architecture (top) and detailed Inception module(bottom) [13].

The I3D structure is shown in Fig. 1, contains 4 convolutional layers, 4 Max Pooling layers, 1 Average pooling layer and 9 Inception modules. A detailed Inception module is also shown in Fig. 2. (bottom). The inception module uses 1×1×1 and 3×3×3 convolutional filters for calculating multi-scale information and also the dimensionality reduction. The I3D model is pretrained on kinetics dataset.

The framework used is PyTorch which helps in the fast implementation of the deep neural networks. It uses CUDA support for enabling GPU (Graphics Processing Unit).

4. EXPERIMENTS AND RESULTS

This section introduces the dataset and then the experimental results will be given.

4.1. DATASET

We evaluated the I3D model on the Word Level American Sign Language (WLASL) Dataset [14], this dataset consists of videos from different websites collected into json files. It also has four subsets, Table. 1 shows the details of the dataset. The ASL Alphabet dataset [14] is used to implement in the Amazon Rekognition service.

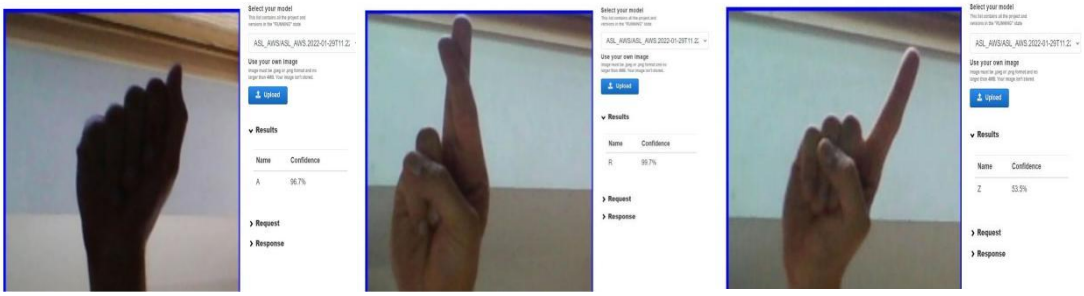
Table 1. Details of the WLASL dataset. #Classes, #Videos denotes the number of classes, videos. #Train, #Test denotes the number of videos taken for training and testing. #Mean denotes average number of videos per class.

Subset	Classes	Videos	Train	Test	Mean
WLASL100	100	2,038	1,780	258	20.4
WLASL300	300	5,118	4,450	668	17.1
WLASL1000	1000	13,174	11,298	1,876	13.2
WLASL2000	2000	21,095	18,216	2,879	10.5

4.2. RESULTS

ISSN PRINT 2319 1775 Online 2320 7876
Research paper **Volume 11, Sp.Iss 5 , 2022**

The ASL Alphabet dataset is used for the implementation Amazon Rekognition which uses a Deep CNN algorithm. The training accuracy obtained in the Amazon Rekognition Custom Labels is 100%. Using Amazon Cloud Front service we can run our model in a website. Fig. 2. shows the detection of signs of characters A,F,R,Z in the cloud front stack, the Confidences obtained for 'A', 'F', 'R', 'Z' is 96.7%, 95.5%, 99.7%, 53.5%.



(a) Detection of Sign 'A' (b) Detection of Sign 'R' (d) Detection of Sign 'Z'

Fig. 2. (a),(b),(c) Character signs detection using Amazon Rekognition.

The prediction of given word sign with the accuracy level which is tested on the I3D trained model is shown in Fig. 3. The model predicted the outputs as "insurance"(a), "magic"(b), "know"(c), "feel"(d) with accuracies 90%, 50%, 82%, 100%.

ISSN PRINT 2319 1775 Online 2320 7876
Research paper Volume 11, Sp.Iss 5 , 2022

Fig. 3. Outputs of word signs video level detection using I3D. These images are taken from the ouput video obtained.

The Subsets WLASL100, 300, 1000, 2000 has shown approximately 89%, 76%, 69%, 59% accuracies accordingly.

CONCLUSION

In this paper we present a sign language detection system for word level signs in videos. The developed sign language detection system detects the signs from the videos with an overall testing accuracy of 66%, which shows an improvement than the other implemented models. The Amazon Rekognition has given a 87~90% accuracy for the images of the same dataset and an average of overall testing accuracy 72~80% for the images outside the dataset has been shown by the service model. Further, If we can increase the capacity of our dataset and the speed of the gpu, and train this model it shows good improvement in terms of accuracy and precision. If the accuracy is improved, we can build a continuous translation system for real time translation.

REFERENCES

[1] Hall, M. L., Hall, W. C., & Caselli, N. K. (2019). Deaf children need language, not (just) speech. First Language, 014272371983410. doi:10.1177/0142723719834102

[2] Ss, Shivashankara& S, Dr.Srinath. (2018). American Sign Language Recognition System: An Optimal Approach. International Journal of Image, Graphics and Signal Processing. 10. 10.5815/ijigsp.2018.08.03.

[3] Suharjito, Anderson, R., Wiryana, F., Ariesta, M. C., & Kusuma, G. P. (2017). Sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on

ISSN PRINT 2319 1775 Online 2320 7876
Research paper **© 2012 IJFANS. All Rights Reserved**, Volume 11, Sp.Iss 5 , 2022

Input-Process-Output. Procedia Computer Science, 116, 441–448. doi:10.1016/j.procs.2017.10.028

[4] Kakde, Manisha &Nakrani, Mahender &Rawate, Amit. (2016). A Review Paper on Sign Language Recognition System For Deaf And Dumb People using Image Processing. International Journal of Engineering Research and. V5. 10.17577/IJERTV5IS031036.

[5] Athira, P. K., Sruthi, C. J., &Lijiya, A. (2019). A Signer Independent Sign Language Recognition with Co-articulation Elimination from Live Videos: An Indian Scenario. Journal of King Saud University - Computer and Information Sciences. doi:10.1016/j.jksuci.2019.05.002

[6] Ankit, O. , Ayush, P. , Shubham, M. , Abhishek, T. ,Dayananda, P. , 2020, Sign Language to Text and Speech Translation in Real Time Using Convolutional Neural Network, INTERNATIONALJOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) NCAIT – 2020 (Volume 8 – Issue 15).

[7] Chiarella, Davide &Bibuli, Marco & Bruzzone, Gabriele & Caccia, Massimo & Ranieri, Andrea &Zereik, Enrica & Marconi, Lucia & Cutugno, Paola. (2015). Gesture-based Language for Underwater Diver-Robot Interaction. 10.1109/OCEANS-Genova.2015.7271710.

[8] Pal, Shantanu & Mukhopadhyay, Subhas & Suryadevara, Nagender. (2021). Development and Progress in Sensors and Technologies for Human Emotion Recognition. Sensors (Basel, Switzerland). 21. 10.3390/s21165554.

[9] Tamura, S., & Kawasaki, S. (1988). Recognition of sign language motion images. Pattern Recognition, 21(4), 343–353. doi:10.1016/0031-3203(88)90048-9

[10] Gupta, B., Shukla, P., & Mittal, A. (2016). K-nearest correlated neighbor classification for Indian sign language gesture recognition using feature fusion. 2016 International Conference on Computer Communication and Informatics (ICCCI). doi:10.1109/iccci.2016.7479951

[11] Rao, G. A., Syamala, K., Kishore, P. V. V., & Sastry, A. S. C. S. (2018). Deep convolutional neural networks for sign language recognition. 2018 Conference on Signal Processing And Communication Engineering Systems (SPACES). doi:10.1109/spaces.2018.8316344

ISSN PRINT 2319 1775 Online 2320 7876
Research paper **Volume 11, Sp.Iss 5 , 2022**

[12] Kumar M. , Gupta P. , Jha R. K. , Bhatia A. , Jha K. and Shah B. K. , Sign Language Alphabet Recognition Using Convolution Neural Network, 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp.1859-1865, doi: 10.1109/ICICCS51141.2021.9432296.

[13] Carreira, J., & Zisserman, A. (2017). Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2017.502

[14] Li, D. , Rodriguez, C. , Yu, X. , and Li, H. , Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison, in Proc. of WACV, 2020, pp. 1459–1469.

[15] Akash, N. , (2018, April). ASL Alphabet , Version 1, Retrieved January 22, 2022 from https://www.kaggle.com/grassknoted/asl-alphabet