Sign Language Detection Using Action Recognition

[1]Vikas G N, [2]Guruprasad V, [3]Siddharth Bavale, [4]Gagan K B, [5]Nikhil S Tengli

[12345]School of Computer Science and Engineering

REVA University, Bangalore, Karnataka, India

[1]vikasgn2@gmail.com,[2]guruprasadv22@gmail.com

[3]sid19sep@gmail.com,[4]kbgagan@gmail.com,[5]nikhil.tengli@reva.edu.in

Abstract

Sign language is a visual language that is used to communicate through hand gestures, hand shape changes, and track information. This is the primary means of communication for the hearing impaired. Despite the fact that action identification in videos has been extensively researched, many algorithms fail to detect actions in real time[3]. Many new methods for dealing with difficult video sequences enhance accuracy and resilience. Recent advances in the application of deep learning to natural language processing and machine translation have left hearing-impaired people, who are part of the population, behind [2]. Here we implement a system which allows this segment of people benefit from these improvements. Transfer learning is employed in our system to speed up the learning process and improve results.

Keywords: Action Recognition, Sign Language

1. INTRODUCTION

ASL and BSL are the world's most widely used sign languages. In the United States, UK and Canada, about 650,000 people use it. We combine ASL and action recognition to analyze the actions performed by humans and respond accordingly. Analysis of video footage is a very expensive task and even the most top of the line machines with very good specifications sometimes have trouble running these tasks [6]. In recent years, deep learning methods have displayed promising results after learning from data obtained by the device.

This concept can be repurposed and used in a lot of different applications. The project done here is using it to detect sign language and make the system respond to the sign language performed by the user accordingly. The software detects sign language in real time . A model is generated based on this for use.

The model helps muted people to communicate with normal people and express themselves. It becomes a lot easier for the normal people to understand muted people.

The feature detection is completed using various contour analysis and have extractions. Thus in the end a tool which can recognize sign language and its other variations.

2. RELATED WORKS

Many approaches and methods have been proposed for Action Recognition. This section enlists the efforts made in the prior:

1. In research paper [1] the authors train neural networks for image classification. Complex images are also trained with MNIST models. This lead to the development of Recurrent Neural Network for the images used. The classification improved to such an extent that even the human eye could not differentiate the classification of the images.

2. In research paper [2] For training, a balanced amount of face and non-facial images are employed. Using a bi-scale CNN 120 that has been auto-stage trained and developed imagenet classification with the help of Binary Convolutional Neural Network. This resulted in a state of art 80% detection rate with just about 50 false positives.

3.In research paper [3] Under each Hierarchical classifier, there are several datasets. Rejection of the class based on the intermediary stage was performed by the authors. The survey done by the authors considered classification techniques such as Decision Tree (DT), Support Vector Machine (SVM), and Fuzzy Classification under the artificial neural networks.

4. In research paper [4] Spectral information is combined with spatial information from a sequential trial method. This lead to SVM Active Learning Approach for Image Classification Using Spatial Information. The results obtained by the authors in this research paper demonstrate the efficacy of regularization in the spatial domain for active learning purposes.

5. In research paper [5] Simply increasing Meta knowledge where local characteristics can be mostly found. This approach followed by the authors gives better classification accuracy. This paved the way to a reduction in the time required for learning and testing process by around more than 30%.

3. OBJECTIVES

The goal is to develop an application which utilizes video groupings for continuous gesture based communication. To foster an application which utilizes activity acknowledgment as a standard. To fabricate a compelling gesture based communication model. To plan an application which utilizations Action Detection. Activity discovery should be fueled by LSTM (Long Short Term Memory Network) layers to construct the model.

4. SYSTEM REQUIREMENTS

CPU : Intel i5 and Ryzen 1st gen +

HDD : Min of 10gb

HD Webcam

OS: Min Win 7 and Above, Linux or MacOS

Python : 3.3 and above

Tensorflow JS, React JS

5. METHODOLOGY

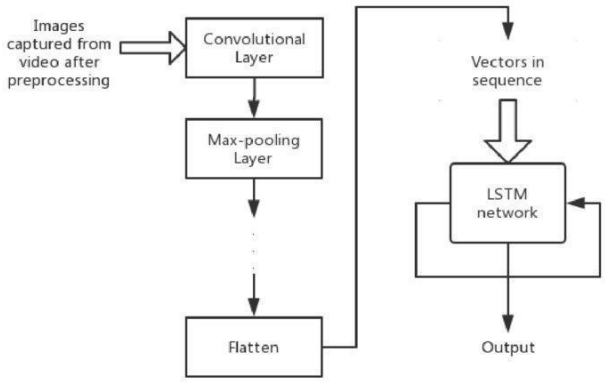

Fig. 1. LSTM model workflow

Live video feed is obtained from the system camera using OpenCV. Keypoints are extracted using MP holistics. Training and testing of data is then performed. Preprocessed data is used to create Labels and Features. Later a deep learning model is Build and Train. Predictions and evaluations are made with the help of Confusion Matrix and Accuracy metrics.

The option to use a keypoint extraction to assemble an arrangement of keypoints with the use of numpy which can then be passed to an activity identification model to perform communication is made possible with the help of OpenCV. As a component of the model creation process the option to use Tensorflow and Keras to construct a Neural network that use LSTM layers to deal with the arrangement of keypoints is performed.

Testing is done to find and rectify errors. In case of poor accuracy the model can be retrained by specifying the appropriate number of epochs. Finally deployment of the model is performed.

6. RESULTS

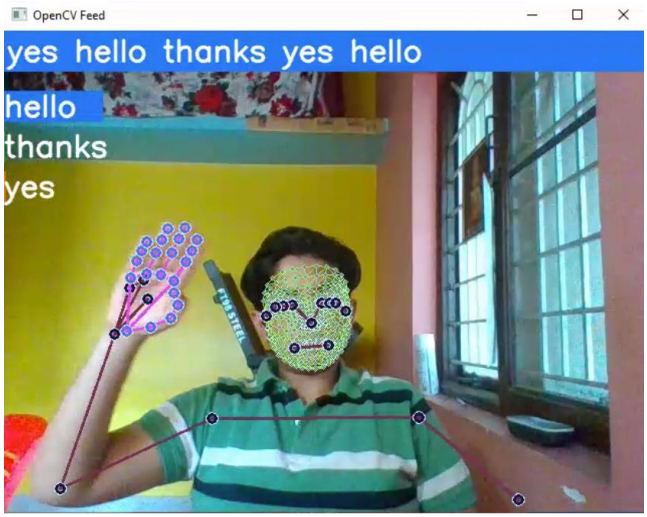

Fig. 2. Action recognition with MediaPipe

In the given image above labeled as **Fig 2** we take 30 video frames to detect a single action performed by the user, thus ensuring higher precision and accuracy compared to single frame action detection model as in **Fig 3**

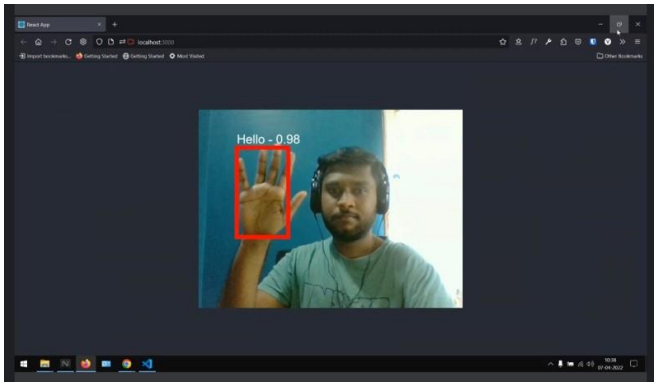

Fig. 2. Action detection with pre-trained tensorflow model

7. CONCLUSION

Computing is a very important tool in providing accessibility for sign language users. Here Action Recognition is being used as a means to detect sign language actions performed in front of a video recording device such as a camera. Many iterations of training may be done to improve the accuracy of the model obtained. The output obtained

is either displayed as a means of communication or further used as input in other applications whichever is deemed necessary

8. REFERENCES

[1] R. U. Shekokar and S. N. Kale, "Deep Learning for Human Action Recognition," 2021 6th International Conference for Convergence in Technology (I2CT), 2021, pp. 1-5, doi: 10.1109/I2CT51068.2021.9418080.

[2] A. L. Nair S. and R. K. Megalingam, "Human Action Recognition: A Review," 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART), 2021, pp. 249-252, doi: 10.1109/SMART52563.2021.9676211.

[3] S. He, "Research of a Sign Language Translation System Based on Deep Learning," 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM), 2019, pp. 392-396, doi: 10.1109/AIAM48774.2019.00083.

[4] N. Zhang, Y. Wang and P. Yu, "A Review of Human Action Recognition in Video," 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS), 2018, pp. 57-62, doi: 10.1109/ICIS.2018.8466415.

[5] S. Das, A. Chaudhary, F. Bremond and M. Thonnat, "Where to Focus on for Human Action Recognition?," 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), 2019, pp. 71-80, doi: 10.1109/WACV.2019.00015.

[6] M. Burić, M. Pobar and M. I. Kos, "An overview of action recognition in videos," 2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2017, pp. 1098-1103, doi: 10.23919/MIPRO.2017.7973588.

[7] Karol Gregor, Ivo Danihelka, Alex Graves, Danilo Jimenez Rezende, Daan Wierstra, DRAW: A Recurrent Neural Network For Image Generation, https://doi.org/10.48550/arXiv.1502.04623

[8] Becky Sue Parton, Sign Language Recognition and Translation: A Multidisciplined Approach From the Field of Artificial Intelligence, The Journal of Deaf Studies and Deaf Education, Volume 11, Issue 1, Winter 2006, Pages 94–101, https://doi.org/10.1093/deafed/enj003

[9] Mohammad Rastegari, Vicente Ordonez, Joseph Redmon, Ali Farhadi "XNOR-Net: ImageNet Classification Using Binary Convolutional Neural Networks" https://doi.org/10.48550/arXiv.1603.05279

[11] Pooja Kamavisdar, S. Saluja, Sonu Agrawal "A Survey on Image Classification Approaches and Techniques" Corpus ID: 212551657

[12] Edoardo Pasolli, Farid Melgani, Yakoub Bazi “Support Vector Machine Active Learning Through Significance Space Construction” DOI:10.1109/LGRS.2010.2083630

[13] Marcin Korytkowski, Leszek Rutkowski, Rafał Scherer. “Fast Image Classification by Boosting Fuzzy Classifiers”

https://doi.org/10.1016/j.ins.2015.08.030

[14] Shreyank N Gowda, Marcus Rohrbach, Laura Sevilla-Lara, “SMART Frame Selection for Action Recognition”, To be published in AAAI-21, Computer Vision and Pattern Recognition, arXiv:2012.10671

[15] H. Jhuang, J. Gall, S. Zuffi, C. Schmid and M. J. Black, "Towards Understanding Action Recognition," 2013 IEEE International Conference on Computer Vision, 2013, pp. 3192-3199, doi: 10.1109/ICCV.2013.396.