NATURAL SENTENCE GENERATION USING SIGN LANGUAGE GESTURES

Mrs. Ancy Y Das , Assistant Professor, Computer Science and Engineering at BGS Institute of Technology, Adichunchanagiri University , ancyydas@bgsit.ac.in ,
BG Nagara , Karnataka

Sushmitha C S, 4th Year, 8th Sem Department of Computer Science and Engineering at BGS Institute of Technology, Adichunchanagiri University cssushmitha38@gmail.com ,

BG Nagara, Karnataka

Abstract: The utilize of sign dialect essentially improves communication for people with hearing disabilities, serving as the essential mode of communication for those who don't utilize verbal dialect. American Sign Dialect (ASL) is famous for its adequacy in encouraging communication among the hearing-impaired. This inventive arrangement leverages innovation to bridge the communication boundary between non-signers and people with hearing disabilities, advancing inclusivity and shared understanding. The proposed venture utilizes progressed innovation to identify ASL signals and interpret them into coherent English sentences, in this manner empowering consistent communication between underwriters and non-signers. By joining MobileNetV2 for sign dialect signal acknowledgment and utilizing Common Dialect Handling (NLP) strategies for sentence era, this activity not as it were distinguishes visual expressions in ASL but too changes them into important English sentences.This combination of computer vision and dialect handling advances holds colossal guarantee in upgrading communication openings for people who are hard of hearing or difficult of hearing, enabling them to lock in more viably with the more extensive community.

Keywords: WLASL, NLP, MobileNetV2, Profound Learning, CNN, T5, TKinter, Sign Dialect.

1. Introduction

Sign language acts as an essential mode of communication for individuals with speech and hearing impairments. However, communicating with non-sign language users can pose challenges. The proposed system will adopt a two-stage approach. In the initial stage, deep learning models will be employed to recognize sign language gestures. Subsequently, in the second stage, NLP techniques such as natural language processing and sentiment analysis will be utilized to generate sentences in natural language based on the recognized gestures.

The system's output will be presented in a user-friendly format, facilitating more efficient communication bridging the gap between sign language users and non-users. In contexts like public services, education, and healthcare, where interaction between sign language users and non-users is crucial, this automated system can mitigate the challenge. While sign language interpreters offer a solution, their services are often costly and not readily available. The objective of this project is to develop a reliable system capable of precisely translating gestures from sign language into spoken language expressions.

By leveraging deep learning and NLP, the system seeks to close the communication divide between sign language proficient individuals and those unfamiliar with sign language. Training the system on standard datasets of sign language gestures and corresponding natural language sentences will ensure precise and reliable translation. Ultimately,

The project endeavors to enrich communication between sign language users and non-signers, elevating the significance of sign language more accessible to a broader audience, including those unfamiliar with it. Through this initiative, the goal is to promote inclusivity and improve communication across diverse communities.

2. LITERATURE SURVEY

The first paper titled "Sign Language Detection using Action Recognition" [1] proposes a deep learning approach to recognize sign language gestures by leveraging action recognition techniques. The authors advocate for enhancing Improving the precision and velocity of sign language recognition systems via the utilization of action recognition models. They argue that such models can effectively capture the temporal dynamics inherent in sign language gestures, ultimately improving system performance. The paper precisely employs an LSTM model for instantaneous sign language detection, utilizing Media Pipe holistic key points for both training and testing phases.

On the other hand, the second paper, "A Comprehensive Survey on Sign Language Recognition: From Traditional Methods to Deep Learning Approaches" [2], provides an extensive overview of SLR. It emphasizes the significance SLR technology. The authors highlight recent advancements which have spurred considerable research interest in SLR. The paper offers a comprehensive survey of SLR techniques, encompassing both traditional methods and deep learning approaches. Additionally, it discusses available datasets and evaluation metrics commonly utilized in SLR research. Moreover, the authors delve into the challenges and characteristics of SLR, including the variability of sign language, the necessity for robustness to lighting and background conditions, and the importance of large and diverse datasets.

While both papers contribute valuable insights to the field of sign language recognition, it's crucial to acknowledge the unique contributions of each study and the broader context in which they operate.

3. Methodology

A. Dataset Description

Our project utilized the WLASL video dataset sourced from Kaggle, which comprises 21,000 sign videos encompassing 2,000 words performed by 100 different signers [13]. The video lengths vary between 0 to 10 seconds, with each video possessing a unique video_id and being linked to its respective gloss through a JSON file. For our study, we focused on a subset consisting of the 200 most commonly used daily conversation words. From each gloss/word category, we selected 5 videos for further processing.

B. Data Pre-processing

It is a crucial step aimed at enhancing the pertinent features necessary for subsequent processing and analysis. This typically involves several tasks such as frame extraction, elimination of sample noise, normalization of frames, and conversion of frames to gray scale to augment data quality and facilitate more efficient feature extraction.

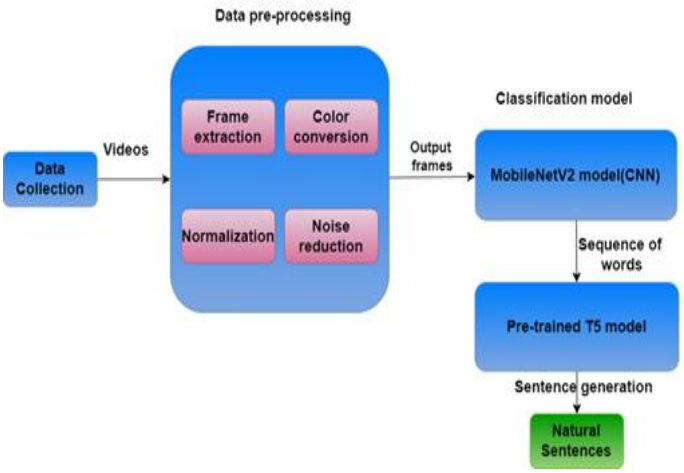

Fig. 1. High-Level Design

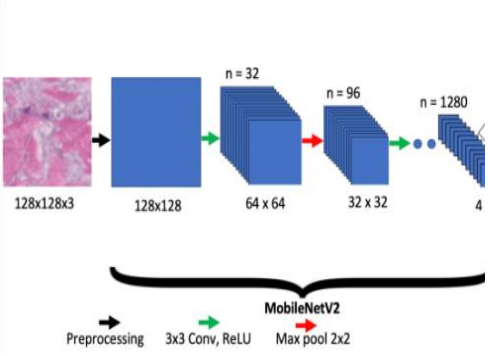

Fig 2. Classification using MobileNetV2

C. Classification model

A pre-trained deep learning model, MobileNet V2, is fine-tuned on an extensive dataset of sign language videos to identify and classify diverse sign language expression. We are using Adam optimizer that keeps each parameter's adaptive learning rate constant enabling it to handle sparse gradients and sparse categorical cross entropy loss function that measures the discrepancy between the actual output and the predicted output" could be error metric evaluating the disparity between observed and forecasted results. The layers of the pre-trained model are frozen so that the weights during training are retained.

D. Sentence Generation

The generation of natural sentences is done by using the 'pipeline' function from the Hugging Face Transformers library. It creates an NLP pipeline for text generation. The pipeline is initialized with a pre-trained T5 (Text-To-Text Transfer Transformer) model and fine-tuned for common text generation tasks. The sequence of predicted words is filtered in such a way that no words are repeated in a list and this list is fed into the NLP pipeline to get a meaningful sentence. The generated sentence gets printed on a user interface.

4. Results

Sign gesture detection: The trained model successfully detects the signs and can store and print the predicted words or labels.

Capture live: The application can capture the live frames and predict the words.

Upload video files from the system: When a video is uploaded from the local system as an input the application is able to predict the words and store the sequence of words in a list, which are then given to the NLP model as an input.
Sentence generation: The trained model is capable of generating a sentence using the sequence of keywords that are given as input.

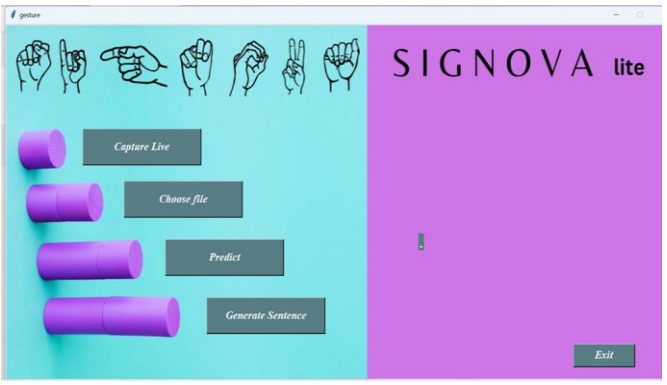

Figure 3. The user interface in the beginning:

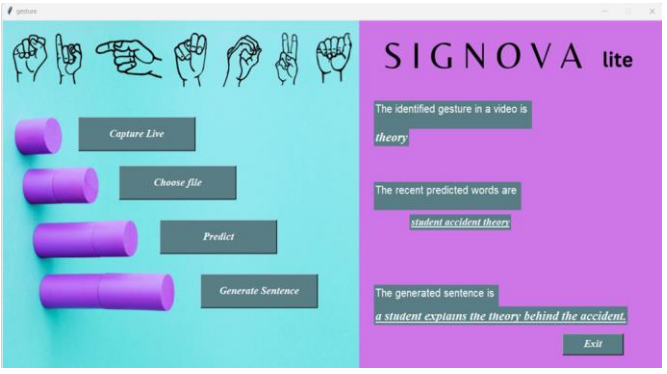

Figure 4: The user interface with the final output:

The proposed model achieved 86% accuracy for sign gesture detection.

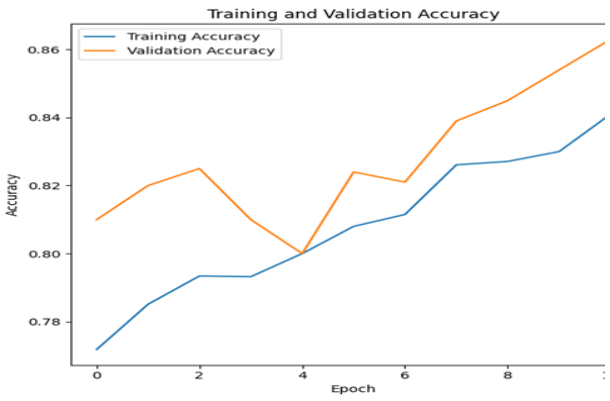

Fig. 5. Training and validation accuracy

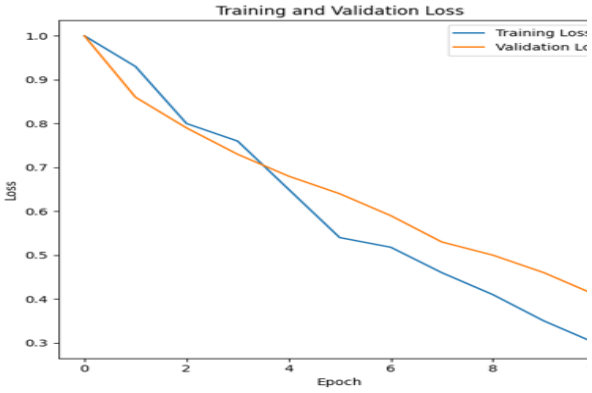

Fig. 6. Training and validation loss

5. Conclusion

The application is created with a user-friendly GUI using the Tkinter toolkit. It can detect various expressions using live capture in real-time and also using a video uploading method. The repeated words are deleted from a sequence of detected words which are then used in a sentence generation model. The final output of the application is a grammatically correct and meaningful sentence.

REFERENCES

[1] V. H. Iyer, U. M. Prakash, A. Vijay, and P. Sathishkumar, "Sign Language Detection using Action Recognition," 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2022.

[2] N. Adaloglou et al., "A Comprehensive Study on Deep Learning-Based Methods for Sign Language Recognition," in Transactions on Multimedia, vol. 24, pp. 1750-1762, 2022.

[3] S. E. Panneer and M. Sornam, "Recent Advances in Sign Language Recognition using Deep Learning Techniques," 2022 6th International Conference on Trends in Electronics and Informatics

[4] Dr.P. Golda Jeyasheeli, N. Indumathi, “Sentence Generation for Indian Sign Language Using NLP” April 2021 Webology 18(Special Issue 01):196-210

[5] Gajjar, Kevin & Agrawal, Aman & Gonsalves, Arran & Singh, Gargi. (2022). Sentence Formation Using NLP on the Basis of American Sign Language. International Journal for Research in Applied Science and Engineering Technology.

[6] Villaplana, Aitana & Martínez-Hinarejos, Carlos-D. (2021). “Generation of Synthetic Sign Language Sentences”. 235-239. 10.21437/IberSPEECH.2021-50. (Conference Paper)

[7] Saad Hassan, Sooyeon Lee, Dimitris Metaxas, Carol Neidle, and Matt Huenerfauth. 2022. Understanding ASL Learners’ Preferences for a Sign Language Recording and Automatic Feedback System to Support Self-Study. In Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '22). Association for Computing Machinery, New York, NY, USA, Article 85, 1–5.

[8] Thakar, Shubham & Shah, Samveg & Shah, Bhavya & Nimkar, Anant. (2022). “Sign Language to Text Conversion in Real Time using Transfer Learning”.

[9] Ronglai Zuo, Fangyun Wei, Brian Mak,“Natural Language-Assisted Sign Language Recognition” Accepted by CVPR 2023.

[10] D. Li, C. R. Opazo, X. Yu and H. Li, "Word-level Deep Sign Language Recognition from Video: A New Large-scale Dataset and Methods Comparison," 2020 Winter Conference on Applications of Computer Vision (WACV), Snowmass, CO, USA, 2020, pp. 1448-1458, doi: 10.1109/WACV45572.2020.9093512.