Conference Paper Title*

Reuben Rouse
Department of Artificial Intelligence
Mukesh Patel School of Engineering
Mumbai, India
reubenjrouse@gmail.com

Lehar Rathore
Department of Artificial Intelligence
Mukesh Patel School of Engineering
Mumbai, India
leharrathore25@gmail.com

Dr. Vaishali Kulkarni
Department of Artificial Intelligence
Mukesh Patel School of Engineering
Mumbai, India
email address or ORCID

Abstract—Indian Sign Language is a vital communication tool for the deaf community in India, characterized by a mix of static and dynamic gestures. However, the limited number of interpreters and the increasing reliance on digital communication have widened the communication gap. This paper proposes a hybrid approach to Continuous Sign Language Recognition (CSLR) and translation, leveraging a sign spotter trained on a large linguistic dataset and a Large Language Model (LLM) to generate spoken language outputs. We achieved 94% accuracy using a ConvLSTM for the sign spotter, which was trained on 6 gestures.

Index Terms—Indian Sign Language, glosses, ConvLSTM, CNN, LSTM

I. INTRODUCTION

Sign language is a visual means of communication that uses hand gestures, facial expressions, and body language to convey meaning, serving as a vital communication tool for those who are deaf or hard of hearing. However, sign language is not universal; different countries have their own distinct sign languages. For example, American Sign Language (ASL) and Indian Sign Language (ISL) use different signs and grammar, making sign language unique to each region.

In India, approximately 63 million people suffer from significant auditory impairment, according to WHO estimates. Despite this, there are only about 339 Indian Sign Language interpreters available. This disparity results in a significant communication gap, as the ratio of deaf individuals to interpreters is exceedingly high. The situation has worsened in the digital age, particularly following the COVID-19 pandemic, which has led to a shift towards online meetings and classes. This transition has exacerbated communication challenges for those reliant on sign language, underscoring the urgent need for more accessible translation solutions.

Indian Sign Language (ISL) is a standardized system of sign language gestures used by the speech-impaired community across India. ISL gestures can be broadly categorized into two types: static and dynamic. Static gestures involve holding a hand in a fixed position without movement and are typically used for representing letters and simple signs. For instance, most alphabetic gestures in ISL are static. In contrast, dynamic gestures incorporate hand movements and are used for more complex signs and expressions. These dynamic gestures make up the majority of ISL signs, reflecting the language's rich expressiveness and nuance.

Sign Language Translation (SLT) is an essential task that focuses on converting sign language into spoken or written language, making it a crucial tool for bridging communication gaps. SLT is often framed as a Neural Machine Translation (NMT) problem, where the goal is to generate coherent spoken or written sentences from sign language videos. However, SLT presents unique challenges compared to traditional text-based NMT due to the continuous nature of sign language videos and the differences in grammar between sign languages and spoken languages.

Sign Language Translation (SLT) is the process of converting sign language into spoken or written language. It is often approached as a Neural Machine Translation (NMT) task, where the goal is to translate sign language videos into coherent spoken or written sentences. However, unlike text-based NMT, SLT involves dealing with continuous sign language videos that are challenging to align and tokenize. Additionally, the grammar structure of sign languages like Indian Sign Language (ISL) differs significantly from English. ISL typically follows a Subject-Object-Verb (SOV) word order, whereas English follows a Subject-Verb-Object (SVO) order. This difference further complicates direct translation.

To address these challenges, the focus is on Continuous Sign Language Recognition (CSLR) rather than isolated sign language recognition (ISLR). CSLR involves recognizing sequences of glosses (a representation of signs) as an intermediate step before translating them into spoken language. This hybrid approach leverages a sign spotter to identify glosses, which are then passed to a Large Language Model (LLM) for generating the final spoken language output. This method bypasses the need for extensive SLT-specific training or datasets.

REFERENCES

[1] O. M. Sincan, N. C. Camgoz, and R. Bowden, Using an LLM to Turn Sign Spottings into Spoken Language Sentences. 2024.

[2] J. Gong, L. G. Foo, Y. He, H. Rahmani, and J. Liu, LLMs are Good Sign Language Translators. 2024.

[3] E. Abraham, A. Nayak and A. Iqbal, "Real-Time Translation of Indian Sign Language using LSTM," 2019 Global Conference for Advancement in Technology (GCAT), Bangalore, India, 2019, pp. 1-5.

[4] Dabwan , Basel & Jadhav, Mukti. (2023). A CNN-LSTM Model for Arabic Sign Language Recognition. 10.2991/978-94-6463-196-8_35.

[5] Jie Huang, Wengang Zhou, Houqiang Li and Weiping Li, "Sign Language Recognition using 3D convolutional neural networks," 2015 IEEE International Conference on Multimedia and Expo (ICME), Turin, 2015, pp. 1-6.