

Multimodal Attention-Based Framework for Sign Language Recognition

Swapnil Ohol¹ and Dr. R. A. Kulkarni²

MTech Student, Department of Computer Engineering¹

Professor, Department of Computer Engineering²

Pune Institute of Computer Technology, Pune, India

Abstract: Sign language recognition (SLR) plays an essential role in enabling effective communication between individuals with hearing impairments and the hearing community. Conventional SLR systems often depend on a single input type, such as RGB frames or skeletal data, which limits their accuracy and adaptability in complex, real-world environments. This dissertation presents a multimodal attention-based framework that integrates RGB video, skeletal joint, and landmark information to comprehensively model the spatial and temporal aspects of signing. The system leverages transformer-based attention mechanisms, graph convolutional networks (GCNs), and recurrent layers to capture both global dependencies and fine-grained gesture dynamics. A dedicated Head-and-Hands Tunnelling Pipeline is employed to emphasize key gesture regions and minimize background interference. Additionally, the framework incorporates dataset refinement and landmark correction to enhance stability during training. Experimental evaluation will be conducted on benchmark datasets such as MS-ASL and WLASL for both isolated and continuous recognition tasks. The overall objective is to design a robust, real-time SLR model that contributes to improved accessibility in educational, healthcare, and social communication contexts.

Keywords: Sign Language Recognition (SLR) , Multimodal Deep Learning , Transformer Networks, Graph Convolutional Networks (GCN)

