Department: CSE Thematic Area:Deep learning and AI

Sign Language Detection by Image Transformer: An AI-Based Framework for Hand Gesture Recognition

, H.K.Borna, S.M.B.Mohammed, S.K.Prova and M.Ahmed

Department of Computer Science and Engineering ,Bangladesh Army International University of Science and Technology

**Corresponding E-mail: mamun.cse@baiust.ac.bd*

Abstract

The Objective of Sign language is a vital communication method for deaf and hard-of-hearing individuals, yet the lack of broad public understanding poses significant communication challenges. This thesis presents the "Sign Language Image Transformer" system, an AI-driven framework combining deep learning, computer vision, and transformer architectures to convert static hand gestures into readable text, bridging communication gaps. The proposed system integrates several advanced AI technologies to build an efficient and adaptable gesture recognition model. MediaPipe is used for accurate hand landmark detection, capturing the shape, orientation, and position of the hand. These landmarks are processed using Convolutional Neural Networks (CNNs) to extract critical visual features. Image and visual transformation techniques are then applied to enhance the system's ability to identify and classify signs precisely. A Transformer Encoder model, incorporating self-attention mechanisms, further processes the features to improve recognition accuracy, even for visually similar gestures. The system's learning process is supported by machine learning and deep learning techniques, ensuring reliable and scalable performance. Experimental implementation demonstrates that the system can accurately recognize a wide range of static sign gestures with high precision and consistency, making it suitable for real-world applications. This research aims to deliver a low-cost, user-friendly, and accurate tool for sign language recognition. By leveraging modern AI methods, the system contributes to making communication more inclusive and accessible for individuals with hearing or speech impairments.

Keywords : Sign language recognition, MediaPipe , CNN , Machine Learning , Deep Learning, Transformer model, Image transformation, Accessibility, Assistive technology.