Real-Time Sign Language Translation to Coherent English Output

Shravan Venkatraman
School of Computer Science and Engineering
Vellore Institute of Technology, Chennai

Jayasankar K S
School of Computer Science and Engineering
Vellore Institute of Technology, Chennai

Pranay Jiljith
School of Computer Science and Engineering
Vellore Institute of Technology, Chennai

Rethik Manoharan
School of Computer Science and Engineering
Vellore Institute of Technology, Chennai

Abstract:

This research paper introduces a unique methodology for achieving real-time sign language translation to coherent English output. The proposed approach utilizes a word-level sign language dataset for word classification, employing the You Only Look Once (YOLO) algorithm. The dataset enables accurate identification and classification of individual signs, laying the foundation for the translation process. Following the classification of words, Natural Language Processing (NLP) techniques are employed to assemble these words into meaningful sentences. This step involves syntactic and semantic analysis to ensure grammatical correctness and coherence in the resultant English sentences. Finally, the synthesized sentences are converted into speech using Text-to-Speech (TTS) APIs, facilitating seamless communication between sign language users and English speakers. The methodology integrates computer vision, machine learning, NLP, and speech synthesis technologies to bridge the gap between sign language and spoken language, enhancing accessibility and inclusivity for the deaf and hard-of-hearing community.

Keywords: Sign language translation, Real-time translation, YOLO, Natural Language Processing, Text-to-Speech API.

Motivation:

The motivation for the proposed work is to bridge the gap between the differently-abled and the rest of the world. The real-time sign language translation methodology introduced in this research paper has numerous potential applications and opportunities for further development. Beyond the immediate impact on enhancing communication and accessibility for the deaf and hard-of-hearing community, this technology can be leveraged in a variety of settings, such as educational institutions, healthcare facilities, and public service organizations. Additionally, as the field of artificial intelligence and natural language processing continues to evolve, there are ample opportunities to refine and expand the capabilities of this methodology, potentially incorporating multimodal input, multilingual support, and even personalized translation features.

Objective:

The real-time sign language translation methodology presented in this research paper has the potential to significantly impact the lives of the deaf and hard-of-hearing community, as well as the broader society. By bridging the communication gap between sign language users and

English speakers, this innovative solution enhances accessibility and promotes inclusivity. It empowers individuals with hearing impairments to engage more effectively in various social, professional, and educational settings, fostering greater integration and participation. Moreover, the adoption of this technology can contribute to the empowerment of the deaf and hard-of-hearing community, providing them with greater autonomy and independence in their daily lives.

Improved Communication: The real-time translation of sign language to coherent English output enables more seamless and effective communication between the deaf and hard-of-hearing community and English speakers.

Enhanced Accessibility: By bridging the language barrier, this methodology enhances accessibility and promotes the inclusion of individuals with hearing impairments in a wide range of social, professional, and educational settings.

Independence: The adoption of this technology empowers the deaf and hard-of-hearing community, providing them with greater autonomy and the ability to navigate their daily lives with increased independence and self-determination.

Problem Statement:

Despite advancements in technology, there remains a significant communication barrier between sign language users and individuals who do not understand sign language, particularly in real-time interactions. This gap inhibits effective communication and access to information for the deaf and hard-of-hearing community, leading to social exclusion and limited participation in various spheres of life.

Traditional methods of sign language interpretation, such as human interpreters or manual translation, are often time-consuming, costly, and not readily available in all contexts. Additionally, existing automatic sign language translation systems face challenges in accurately recognizing and interpreting sign gestures, leading to errors and misunderstandings in translation.

The lack of efficient and real-time translation systems hampers the ability of sign language users to engage in spontaneous conversations, participate fully in educational settings, access healthcare services, and interact with digital content. This communication barrier exacerbates the social isolation and marginalization experienced by the deaf and hard-of-hearing community, limiting their opportunities for personal and professional development.

Therefore, there is a pressing need for a robust and reliable methodology that can accurately translate sign language gestures into coherent English output in real-time. Such a system would facilitate seamless communication between sign language users and English speakers, breaking down barriers to social interaction, education, employment, and access to essential services. By addressing this need, we aim to promote inclusivity, accessibility, and equal participation for individuals with hearing impairments in society.

Introduction:

In the intricate tapestry of human health, few organs command as much attention and reverence as the heart. Its ceaseless rhythm is not only a testament to life's enduring vitality but also a poignant reminder of its fragility. Despite its paramount importance, the heart is vulnerable to an array of maladies, ranging from the benign to the life-threatening. Yet, amidst this complexity lies a beacon of hope - early detection. Recognizing the signs of heart problems before they escalate can be instrumental in averting catastrophe and preserving lives. This is where Cardio Guard steps in - an innovative solution poised to revolutionize the landscape of cardiac health.

Smartwatches equipped with various health sensors and features can potentially help detect several cardiovascular problems or conditions. These include atrial fibrillation (AFib), tachycardia, bradycardia, hypertension, coronary artery disease (CAD), heart failure, ischemic heart disease, and valvular heart disease. Leveraging the capabilities of smartwatches, Cardio Guard aims to empower individuals with early detection capabilities, enabling proactive management of cardiovascular health and ultimately improving overall well-being.

Background Study – Related Works:

- **[1] Visual Analysis of Humans – Sign Language Recognition: A Review:**

The paper provides an overview of sign-language recognition (SLR), including discussions on sign linguistics, data types, feature extraction, manual and non-manual aspects classification, and combining sign classification results. It explores tracking and non-tracking viewpoints for classifying manual aspects of sign, as well as approaches to non-manual aspects. The paper also discusses techniques for combining sign classification into full SLR, drawing parallels with speech recognition techniques, and addressing signer independence and modality combination. Comprehensive overview of SLR from linguistic and technical perspectives. Addresses challenges in feature extraction and classification. Discusses advancements towards continuous sign recognition and signer independence. Highlights the importance of adapting to larger, noisy datasets.

• [2] Australian sign language recognition – Machine Vision and Applications:

The paper presents an automatic Australian sign language (Auslan) recognition system that employs tracking of face and hands, feature extraction, and Hidden Markov Models (HMMs) for recognition. Tracking utilizes simple geometrical features between current and previous frames. The system addresses occlusion by detecting contours using motion cues and the snake algorithm. Features invariant to scaling, rotation, and signing speed are used for recognition, capturing geometrical positioning, shapes, and motion directions. HMMs are utilized for Auslan phrase recognition. High recognition rates achieved at both sentence and word levels. Robustness to occlusion and variations in signing speed. Utilization of HMMs for modeling sign phrases. Invariance to scaling and rotation enhances system performance.

• [3] Sign Language Recognition Using Convolutional Neural Networks

The paper proposes a sign language recognition system using the Microsoft Kinect, convolutional neural networks (CNNs), and GPU acceleration. CNNs automate feature construction, replacing handcrafted features. The system achieves high accuracy in recognizing 20 Italian gestures, with cross-validation accuracy of 91.7%. It also performs well in the

ChaLearn 2014 Looking at People gesture spotting competition, achieving a mean Jaccard Index of 0.789. Utilization of CNNs for automated feature construction. High recognition accuracy achieved across Italian gestures. Generalization capability demonstrated on unseen users and environments. Success in gesture spotting competitions indicates system effectiveness.

Proposed Work:

At the core of this methodology is a comprehensive word-level sign language dataset, which serves as the foundation for the sign classification process. This dataset enables the accurate identification and classification of individual signs, a crucial step in the translation pipeline. By employing the powerful YOLO (You Only Look Once) algorithm [4], the system is able to precisely detect and categorize the various signs, laying the groundwork for the subsequent natural language processing and speech synthesis stages.

1. **Precise Sign Identification**:

   The word-level sign language dataset allows for the accurate detection and classification of individual signs, laying the foundation for the translation process.

2. **Robust Classification:**

   The incorporation of the YOLO algorithm, a state-of-the-art object detection model, ensures reliable and efficient sign recognition, a crucial step towards seamless translation.

3. **Translation Foundation**:

   The accurate classification of individual signs serves as the cornerstone for the subsequent natural language processing and speech synthesis stages, enabling the transformation of sign language into coherent English output.

The final step in the real-time sign language translation process involves the conversion of the synthesized English sentences into spoken output. By leveraging advanced Text-to-Speech (TTS) [5] APIs, the methodology seamlessly transforms the written English into natural-sounding speech, allowing for effortless communication between sign language users and English speakers. This integration of speech synthesis technology ensures that the translation process is complete, enabling a smooth and intuitive exchange of information and ideas.

Results

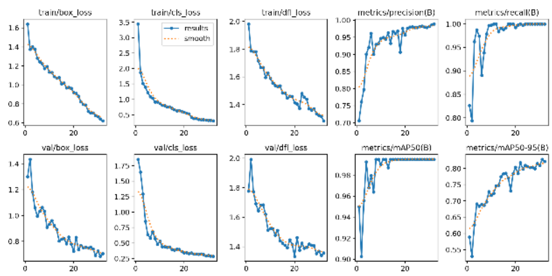

Figure 1: YOLO Training Results

The real-time sign language translation methodology presented in this research paper represents a significant step forward in enhancing communication and accessibility for the deaf and hard-of-hearing community. By seamlessly bridging the gap between sign language and spoken language, this innovative approach empowers individuals with hearing impairments and promotes inclusive interactions. Through the integration of cutting-edge technologies, including computer vision, machine learning, natural language processing, and speech synthesis, this methodology has the potential to transform the way we facilitate communication and foster a more inclusive society. As the field of artificial intelligence and language technology continues to evolve, the impact of this groundbreaking work will only continue to grow, paving the way for a future where language barriers are effortlessly overcome, and communication is truly universal.

References

[1] Cooper, H., Holt, B., Bowden, R. (2011). Sign Language Recognition. In: Moeslund, T., Hilton, A., Krüger, V., Sigal, L. (eds) Visual Analysis of Humans. Springer, London. https://doi.org/10.1007/978-0-85729-997-0_27

[2] Holden, E.-J., Owens, R., & Lee, G. (2005). Australian Sign Language Recognition. Machine Vision and Applications, 16(5), 312-320. https://doi.org/10.1007/s00138-005-0003-1

[3] Pigou, Lionel & Dieleman, Sander & Kindermans, Pieter-Jan & Schrauwen, Benjamin. (2015). Sign Language Recognition Using Convolutional Neural Networks. 8925. 572-578. 10.1007/978-3-319-16178-5_40.

[4] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788)..

[5] Nwakanma, Cosmas & Oluigbo, Ikenna & Izunna, Okpala. (2014). Text – To – Speech Synthesis (TTS). 2. 154-163.