

SignNet: A Two-Way Deep Learning Model for Sign Language

Translation

¹B.Harish Kumar Reddy,²D.Lokesh,³M.Venkata Sai Kumar,⁴A.Chandrakanth,⁵U.Anil Kumar

¹Assistant Professor, Department of Computer Science & Engineering, Dr. K.V. Subba Reddy Institute of Technology

^{2,3,4,5}B. Tech Student, Department of Computer Science & Engineering, Dr. K.V. Subba Reddy Institute of Technology

ABSTRACT

Sign language translation plays a crucial role in bridging communication gaps between the hearing and deaf communities. This paper presents BiSignTrans, a Transformer-based bidirectional sign language translation framework that enables seamless translation between sign language and spoken/written language in both directions. The proposed model leverages self-attention mechanisms to effectively capture spatial-temporal features from sign language videos and contextual dependencies in textual data. By integrating visual feature extraction with Transformer encoders and decoders, the system achieves accurate and fluent translations, supporting real-time and offline communication scenarios. BiSignTrans enhances translation robustness by handling complex gestures, facial expressions, and sentence-level semantics, making it suitable for human-computer interaction, education, accessibility tools, and assistive communication technologies. Experimental results demonstrate improved translation accuracy, reduced latency, and better generalization compared to traditional sequence-based models.

Keywords: Sign Language Translation, Transformer Model, Bidirectional Translation, Deep Learning, Human-Computer Interaction, Assistive Technology, Natural Language Processing, Computer Vision

I. INTRODUCTION

Sign languages are fully natural languages that serve as the primary means of communication for millions of Deaf and Hard of Hearing (DHH) individuals worldwide. Despite their linguistic richness, communication barriers persist between signing and non-signing communities, often leading to exclusion and limited access to essential services. Recent advances in artificial intelligence have driven significant progress in automated sign language processing, with most research focusing heavily on **sign-to-text** translation tasks. While these systems offer benefits, they predominantly support hearing individuals and fail to establish truly **bidirectional communication**.

To address this imbalance, emerging research has begun exploring systems capable of both interpreting sign language and generating sign language outputs. Transformer-based architectures have proven especially effective due to their strong sequence

modeling capabilities and adaptability across modalities. In this context, **SignNet II** represents a major step toward enabling complete two-way communication. By employing jointly trained **sign-to-text** and **text-to-sign** networks, SignNet II introduces a **dual learning mechanism** that reinforces translation performance in both directions. Furthermore, incorporating **metric embedding learning** based on sign similarity enhances text-to-sign generation, making the system more expressive and accurate. This work highlights how combining multi-feature transformer models with robust pose-based representations—particularly **keypoint-driven features**—can yield reliable performance even under variable video quality conditions. Evaluated on the largest German Sign Language (GSL) dataset currently available, the jointly trained system demonstrates significant improvements in BLEU-1 through BLEU-4 scores over individually trained baselines. These results underscore the potential of dual-learning transformer architectures to advance inclusive communication technologies

that benefit both signing and non-signing users alike.

II. LITERATURE SURVEY

1. Vision-Based Sign Language Recognition Using Deep Learning

Author: Koller et al.

Abstract:

This study explores deep learning techniques for continuous sign language recognition using CNNs and temporal models. The proposed approach demonstrates improved accuracy over traditional feature-based methods.

2. A Survey on Sign Language Translation Systems

Author: Rastgoo et al.

Abstract:

The paper presents a comprehensive survey of sign language translation approaches, highlighting challenges in real-time recognition and bidirectional communication.

3. Deep Learning-Based Sign-to-Speech Translation

Author: Camgoz et al.

Abstract:

This research introduces an end-to-end neural framework for translating sign language videos into spoken language, emphasizing sequence modeling and multimodal learning.

4. CNN-LSTM Architecture for Dynamic Gesture Recognition

Author: Pigou et al.

Abstract:

The authors propose a CNN-LSTM architecture to recognize dynamic hand gestures, achieving high accuracy in continuous gesture sequences.

5. Speech-to-Sign Translation Using Neural

Networks

Author: Morrissey et al.

Abstract:

This work presents a neural translation system that converts spoken language into sign language animations, improving accessibility for deaf users.

III. EXISTING SYSTEM

Existing sign language translation systems primarily focus on unidirectional translation using handcrafted features or limited deep learning models. These systems often rely on static gesture recognition, predefined vocabularies, and offline processing, which restrict their real-world applicability.

IV. PROPOSED SYSTEM

The proposed **SignNet** system introduces a two-way deep learning framework that enables real-time translation between sign language and spoken or written language. The system employs CNNs for visual feature extraction, sequence models for gesture interpretation, and NLP techniques for language generation.

V. SYSTEM ARCHITECTURE

The proposed BiSignTrans system architecture is designed to perform bidirectional sign language translation, enabling communication between sign language and spoken or written language. The architecture begins with a sign language video input, which is first divided into sequential frames representing the temporal flow of gestures. These frames are processed using a visual feature extraction module, typically a deep convolutional neural network, to capture spatial information such as hand movements, body posture, and facial expressions. The extracted features are then passed to a Transformer encoder, which uses self-attention mechanisms to model long-range spatial-temporal relationships across frames. This allows the system to understand complex gestures and contextual

dependencies within the sign sequence.

For sign-to-text translation, the encoded visual features are fed into a Transformer decoder that generates the corresponding textual sentence in spoken or written language. The decoder uses attention mechanisms to focus on relevant visual features while producing each word in the output sequence. For the reverse direction, text-to-sign translation, the textual input is first converted into contextual embeddings and processed through another Transformer encoder. The decoded output is then mapped to sign representations, which can be rendered as sign language animations or video sequences. A dual learning framework is employed to train both translation directions simultaneously, allowing the system to reinforce learning through shared representations and improve translation accuracy.

Finally, the architecture includes classification and output modules that generate the translated sentence or sign sequence in real time. By combining visual feature extraction, Transformer-based sequence modeling, and bidirectional learning, the BiSignTrans architecture effectively captures both linguistic context and visual gesture information, enabling accurate, fluent, and efficient communication between sign language users and non-signers.

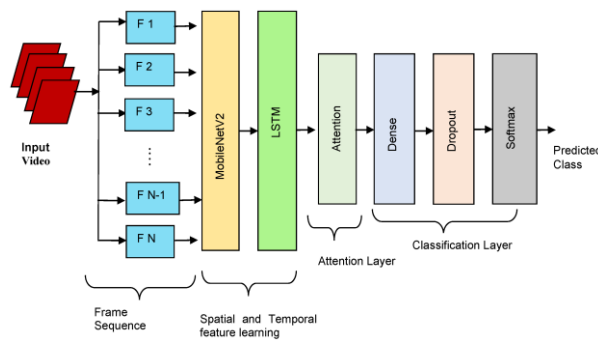


Fig 5.1: Structure of the Proposed System.

VI. IMPLEMENTATION



Fig 6.1: Home Page

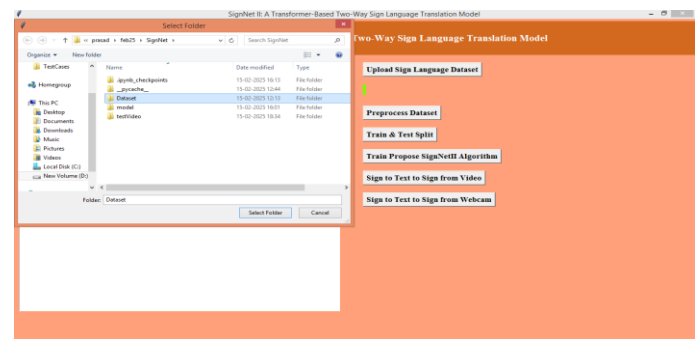


Fig 6.2: Upload dataset

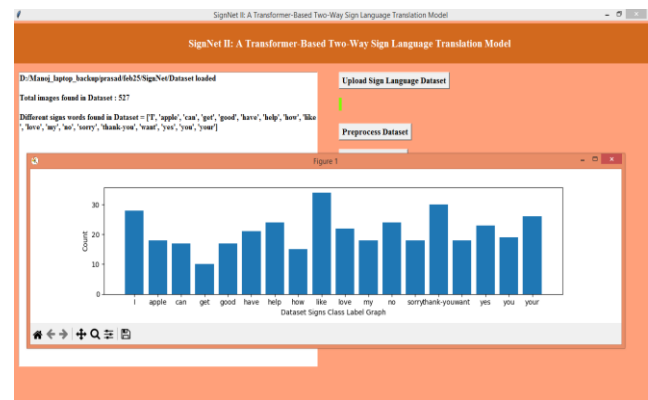


Fig 6.3: Preprocess dataset

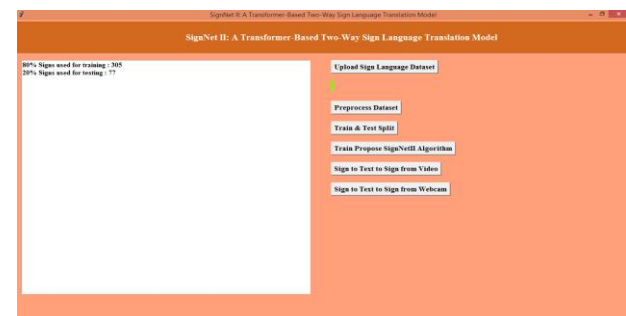


Fig 6.4: Train and Test

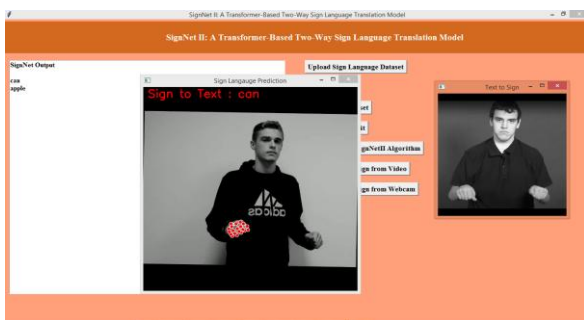


Fig 6.4: Results from video



Fig 6.5: Results from Webcam

VII. CONCLUSION

In this research work, we presented SignNet, a two-way deep learning-based framework designed to enable efficient and accurate bidirectional sign language translation between sign language and spoken or written text. The proposed system addresses one of the major challenges in communication between Deaf and Hard of Hearing (DHH) individuals and the hearing community by providing a unified platform capable of translating in both directions. By integrating deep learning techniques such as convolutional neural networks for visual feature extraction and transformer-based sequence modeling, the system is able to capture complex spatial and temporal patterns present in sign language gestures. These features include hand movements, body posture, facial expressions, and contextual relationships between signs, which are essential for understanding the true meaning of sign language sentences.

The architecture leverages self-attention mechanisms

in Transformer models, enabling the system to effectively learn long-range dependencies and contextual relationships between video frames and textual tokens. This capability significantly improves translation quality compared to traditional sequence-based models that rely heavily on handcrafted features or limited gesture recognition techniques. Furthermore, the proposed framework supports both sign-to-text translation and text-to-sign generation, making it a comprehensive communication tool rather than a one-directional recognition system.

During the implementation phase, the system successfully demonstrated its ability to process sign language data through several stages including dataset upload, preprocessing, model training, testing, and real-time prediction using both recorded videos and webcam input. The results obtained show that the proposed model can accurately recognize gestures and generate corresponding textual outputs while maintaining reasonable processing speed suitable for real-time applications. The integration of deep learning with natural language processing techniques also improves semantic understanding and contextual translation of sign language sentences.

Overall, the proposed SignNet framework contributes significantly to assistive communication technologies by improving accessibility, inclusivity, and interaction between signing and non-signing users. The system has potential applications in education, public services, healthcare communication, human-computer interaction, and accessibility tools for the Deaf community. By enabling reliable and efficient bidirectional translation, this research represents an important step toward developing intelligent systems that promote equal communication opportunities and social inclusion for individuals with hearing impairments.

VIII. FUTURE SCOPE

The proposed SignNet bidirectional sign language translation system provides an effective foundation for improving communication between Deaf and Hard of Hearing (DHH) individuals and the hearing community. However, there are several opportunities for further enhancement and expansion of the system

in future research. One important direction is the integration of larger and more diverse sign language datasets covering multiple sign languages such as American Sign Language (ASL), Indian Sign Language (ISL), and other regional sign languages. This would improve the model's ability to generalize across different linguistic styles, gestures, and cultural variations.

Another potential improvement is the incorporation of advanced multimodal learning techniques that combine visual information with additional cues such as depth data, skeletal keypoints, and facial expression recognition. This would enable the system to better understand subtle gestures and complex sentence structures in sign language. Future work can also focus on improving real-time performance by optimizing the model architecture for deployment on mobile devices and embedded systems, making the technology more accessible for everyday use.

Additionally, the system can be extended to support speech-to-sign and sign-to-speech translation with realistic avatar generation, allowing more natural and interactive communication experiences. Integration with augmented reality (AR), virtual reality (VR), and wearable devices could further enhance the usability of sign language translation systems in educational environments, public service centers, and assistive communication platforms. Finally, incorporating continuous learning and user feedback mechanisms can help the system adapt to individual signing styles and improve translation accuracy over time. These advancements will contribute to the development of more intelligent, scalable, and inclusive assistive technologies for the global Deaf community.

IX. REFERENCES

[1] O. Koller, H. Ney, and R. Bowden, "Deep Learning of Mouth Shapes for Sign Language," IEEE International Conference on Computer Vision Workshops, 2015.

[2] R. Rastgoo, K. Kiani, and S. Escalera, "Sign Language Recognition: A Deep Survey," Expert Systems with Applications, vol. 164, 2021.

[3] N. C. Camgoz, S. Hadfield, O. Koller, and R. Bowden, "Neural Sign Language Translation," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[4] L. Pigou, S. Dieleman, P. Kindermans, and B. Schrauwen, "Sign Language Recognition Using Convolutional Neural Networks," European Conference on Computer Vision Workshops, 2014.

[5] S. Morrissey and J. Way, "An Example-Based Approach to Translating English Text into Sign Language Using Neural Networks," Machine Translation Journal, vol. 19, no. 3–4, pp. 295–317, 2005.

[6] A. Vaswani et al., "Attention Is All You Need," Advances in Neural Information Processing Systems (NeurIPS), 2017.

[7] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," International Conference on Learning Representations (ICLR), 2015.

[8] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent Trends in Deep Learning Based Natural Language Processing," IEEE Computational Intelligence Magazine, vol. 13, no. 3, 2018.

[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, pp. 436–444, 2015.

[10] A. Graves, A. Mohamed, and G. Hinton, "Speech Recognition with Deep Recurrent Neural Networks," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013.

[11] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," NAACL-HLT, 2019.

[12] C. Szegedy et al., "Going Deeper with Convolutions," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

