Enhanced Sign Language Translation between American Sign Language (ASL) and Indian Sign Language (ISL) Using LLMs

Malay Kumar, S. Sarvajit Visagan, Tanish Sarang Mahajan, and Anisha Natarajan

Abstract—We have come up with a research that hopes to provide a bridge between the users of American Sign Language and the users of spoken language and Indian Sign Language (ISL). The research enabled us to create a novel framework that we have developed for Learner Systems. Leveraging art of Large models to create key features including: - Real-time translation between these two sign languages in an efficient manner. Making LLM's capability available for seamless translations to ISL. Here is the full study showing its implementation in this paper. The core of the system is a sophisticated pipeline that begins with reclassification and recognition of ASL gestures based on a strong Random Forest Classifier. By recognizing the ASL, it is translated into text which can be more easily processed. Highly evolved natural language NLP (Natural Language Processing) techniques come in handy as they play a role in our LLM integration where you then use LLMs to be able to convert the ASL text to ISL which provides you with the intent of sentence or phrase. The final step is to synthesize the translated text back into ISL gestures, creating an end-to-end translation experience using RIFE-Net. This framework is tasked with key challenges such as automatically dealing with gesture variability and overcoming the linguistic differences between ASL and ISL. By automating the translation process, we hope to vastly improve accessibility for sign language users. No longer will the communication gap between ASL and ISL create barriers; this totally cool innovation aims to bring our communities closer together. And we believe, with full confidence in our framework, that we're able to apply the same principles across a wide variety of sign language dialects.

Index Terms—Sign Language Recognition, ASL to ISL Translation, Large Language Models (LLMs), Natural Language Processing (NLP), Random Forest Classifier, Gesture Reclassification, Text-to-Gesture Synthesis, RIFE-Net, Real-Time Translation, Sign Language Variability, Linguistic Adaptation, Assistive Technology, Cross-Linguistic Framework, Sign Language Dialects, Accessibility and Inclusivity

I. INTRODUCTION

THE communication gap between the users of American Sign Language (ASL) and Indian Sign Language (ISL) is a significant challenge to intercultural interaction and accessibility in the deaf community. Sign languages are indispensable weapons of expression for deaf individuals, but the lack of interoperability between ASL and ISL limits smooth communication across various linguistic and cultural frontiers. This challenge can be addressed through innovative strategies that leverage advances in machine translation technology and contemporary deep learning frameworks. The central goal of this research work is to build a holistic framework for machine translation that can efficiently and conveniently translate ASL gestures into ISL gestures, hence facilitating communication

between ASL users and ISL users. Toward this objective, we hereby propose an approach that combines image recognition techniques with advanced language processing algorithms governed by Large Language Models. Instead of using traditional CNN- based recognition techniques, LLM-driven techniques are used to decode ASL gestures and, in turn, translate directly into meaningful textual representations. From this step of converting ASL gestures to text, we establish an intermediate that helps use advanced LLM-based techniques for machine translation. Now, we can translate the recognized English text into ISL gestures by preserving linguistic aspects and cultural context. The application of LLMs for translation will help make the process more accurate, context-sensitive, and adaptable and, thus, would be used to bridge between ASL and ISL in a nearly seamless manner.

A. Leveraging Deep Learning for Gesture Recognition and Translation

Sign language communication involves complex expressions that carry a certain range of linguistic and cultural nuances in communication. Conventional sign language understanding and translation techniques rely mostly on fixed data sets and prebuilt models. Such approaches find it difficult to adapt to the dynamic nature of sign languages. In the paper, the constraints were overcome using Random Forest Classifiers for effective gesture recognition followed by Large Language Models to assist in context-aware translation. The proposed framework, therefore, would represent a breakthrough with the integration of real-time processing capabilities and cultural contextualization.

The significance of this work is that it introduces an intermediate text-based representation that forms the connecting link between recognition and synthesis of gestures. This linguistic intermediate allows for not only a more literal translation but also allows the tailoring of the translation method in such a way as to keep its intent and cultural nuances of the original expression. This model would become especially important to the reduction of the linguistic differences between ASL and ISL, such as differing grammatical structures, word orders, and contextual expressions. Further, generating ISL gestures from translated text with the assistance of RIFE-Net produces a smooth and natural gesture presentation. Such a robust ability of RIFE-Net to effectively handle high variability within gesture sequences leads it to accurately reproduce ISL gestures even while making forward-looking predictions of potentially

complex or subtle translations. With advanced recognition, translation, and synthesis modules integrated, the framework places it at the forefront of a cutting-edge tool in the domain of sign language translation.

B. Towards a Multi-dialectal Future

Though designed specifically for ASL and ISL, underpinning principles and methodologies are adaptable to support more than those targeted sign language dialects. By adapting the system to use a variety of datasets, it follows that it is viable to address the diversity in sign languages globally, thereby paving the course for universal sign language interoperability. Moreover, this adaptability brings out the scale in using LLMs for translation purposes as it opens up broader applications in assistive technologies.

This research promises to work out an integrated approach combining gesture recognition, natural language processing, and synthesis of gestures to bridge the overall communication gap so present in sign language communities. This is an advance both in the technical and importance toward furthering the aspects of inclusivity and exchange across the world's deaf community.

II. LITERATURE SURVEY

Recognition and translation of sign languages have attracted a lot of focus in the recent past because of their pivotal role in improving communication between deaf and mute people. Presenting an overview of the literature in this field, this literature review identifies methodologies, techniques, and challenges in sign language recognition and translation systems.

Rao et al. (2023) introduced a system of how people with many spoken languages utter sign language using the Natural Language Toolkit (NLTK) and argued that a critical factor in sign language translation is linguistic processing [1]. A popular review carried out by Al-Qureshi et al. (2023) focused on the DP of recognizing sign languages through deep learning approaches, including challenges and recent milestones [2]. Salian et al. (2017) also presented a system for sign language recognition to set the background for subsequent work in this domain [3].

In this Article, Mubashira and James 2020 proposed a Transformer Network for video-to-text translation proving the transformers effective in sign language [4]. Halder and Tayade (2021) in a recent study that designed a real-time vernacular sign language recognition system using media pipe and machine learning propose the possibility of real-time applications in this field [5]. Also, in Korean sign language recognition, Shin et al. (2023) proposed a Transformer-Based Deep Neural Network and demonstrated how deep learning can be applied for sign language recognition irrespective of the type of sign language [6].

Saleem et al. (2023) provided an outstanding, machine learning-based, full-duplex sign language communication system capable of translating multiple sign languages; the problem of sign language communication technology for the deaf community has received continued attention [35]. Wang et al. (2022) proposed an improved 3D-Res-Net sign language

recognition algorithm with clearer hand features and improved the state of the art of gesture recognition [7].

In their paper, Sahoo et al. (2014) gave a detailed account of the recent approaches to recognizing sign languages, and major issues [8]. Pathan et al. Pathan, Muhammad, and Zhang (2023) developed a fusion approach for sign language recognition from both image and hand landmarks using a multi-headed convolutional neural network demonstrating how multiple-modal can enhance gesture recognition systems [9].

Shenoy et al. (2018) have proposed real-time ISL recognition using grid-based feature extraction and machine learning to address the requirements of the Indian deaf and Hard of hearing community [10]. Sharma and Singh (2022) investigated NLP in the general mapping of speech to ISL with insights towards a hybrid way of sign language translation [11].

In addition, while Badhe and Kulkarni (2015) developed an Indian sign language translator based on gesture recognition algorithms, they pointed out that feature extraction and motion detection remain decisive factors in designing sign language recognition systems [12].

Furthermore, Jia Gong (2024), the author was motivated by the remarkable translation performance of LLMs and proposed a method to incorporate off-the-shelf LLMs to address complex Sign Language Translation Tasks [38]. ZhiGang Chen (2024) thought exploring the gloss-free methods is much more crucial because it will greatly decrease the annotation time and promote the dependence of more accurate and universal sign language translation systems [33]. Sen Fang then unveiled SIGNLLM which is the first large-scale multilingual SLP model created using the prompt2 sign dataset which generates the skeletal poses of sign language from text or prompt for eight languages [34].

III. PROPOSED METHODOLOGY

This research presents a novel methodology to bridge the communication gap between American Sign Language (ASL) and Indian Sign Language (ISL) through an automated translation system enhanced using Large language models (LLMs). The proposed approach comprises three interconnected phases: The sign language recognition phase that utilizes a hybrid ensemble model, the Recognized Text Correction phase through language model enhancement, and the Video synthesis phase with motion smoothing. Each phase has been meticulously designed to address specific challenges in cross-sign-language translation while maintaining semantic accuracy and natural gesture flow.

A. Sign Language Recognition

The initial phase employs a hybrid ensemble approach combining the complementary strengths of a Random Forest Classifier (RFC) and a Convolutional Neural Network (CNN). This RFC+CNN model architecture allows robust feature extraction and classification of the sign language and provides better prediction accuracy.

1) Random Forest Classifier Model: The Random Forest Classifier component focuses on the skeletal approach which sets a baseline for sign language recognition, it uses the hand-tracking capabilities of the mediapipe framework for feature extraction.

Dataset Preparation: For the above model we prepare an image dataset with a sample size of 2800 images (28 distinct classes * 100 images)

- 26 alphabetic characters (100 images per character).
- Additional classes for 'DELETE' and 'SPACE' operations.
- A total of 28 distinct classes.

TABLE I
DATASET DISTRIBUTION AND ACQUISITION PARAMETERS

Parameter	Specification	Justification	
Image Reso-	32×32 pixels	Optimal balance between	
lution		detail preservation and	
		computational efficiency	
Lighting	Ambient, Direct, Shadow Ensures robustness		
Conditions		various lighting scenarios	
Background	Background clutter, ob-	Improves model general-	
Variation	jects	ization	
Camera An-	0°, ±15°, ±30°	Accounts for natural vari-	
gles		ation in hand positioning	
Crop Focus	Hand region	Eliminates irrelevant	
		background information	

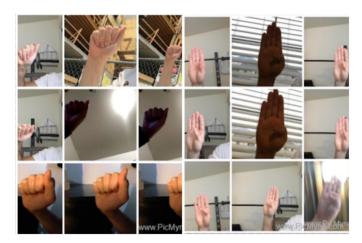


Fig. 1. American Sign Language dataset sample

Feature extraction: With the mediapipe framework we extract 42 prominent landmark points (L) placed on the hand which serve as the features for training the model.

$$L = \{(x_i, y_i, z_i) \mid i \in [1, 42]\}$$

The coordinates obtained are then processed for normalization using StandardScaler.

$$x_{\text{normalized}} = \frac{x - \mu_x}{\sigma_x}$$

where μ_x and σ_x represent the mean and standard deviation of the feature distribution respectively.

Model Architecture and Training: We use the RandomForestClassifer (RFC) model that employs an

TABLE II RFC HYPERPARAMETER CONFIGURATION

Parameter	Search Space	Best Value	Impact
n_estimators	[100, 200, 300]	200	Balances complexity and accuracy
max_depth	[None, 10, 20, 30]	20	Controls overfitting
min_samples_split	[2, 5, 10]	5	Ensures robust node splitting
min_samples_leaf	[1, 2, 4]	2	Prediction stability
bootstrap	[True, False]	True	Enables varied tree construction

optimized through the GridSearch Algorithm.

The model's decision function for class prediction can be expressed as:

$$\hat{y} = \text{mode } \hat{y}t(x)t = 1^T$$

where $\hat{y}_t(x)$ represents the prediction of the t-th tree, and T is the total number of trees.

2) Convolutional Neural Network Model: For the convolutional neural network (CNN) component we use silhouette images providing complementary visual feature analysis to the RFC's landmark-based approach.

Silhouette Images: It is a type of image where a dark image of a subject against a lighter background, usually showing the subject's profile.

Using silhouette images enables a focus on the intrinsic structure of the hand pose, effectively eliminating background noise, lighting variations, and other environmental influences. This approach isolates the key features of the hand's position and configuration, providing a clearer and more accurate representation for training the model.

Dataset Preparation: For the CNN model we generate silhouette images of the American sign language (ASL) with a sample size of 16200 images (27 distinct classes * 600 images).

- 26 alphabetic characters (600 images per character)
- Additional class for "BLANK" image which serves as ground truth.
- Augmentation of images:
 - Horizontal flipping: $I_{flip}(x,y) = I(w-x,y)$
 - Brightness variation: $I_{bright}(x,y) = \alpha I(x,y)$, where $\alpha \in [0.8, 1.2]$
 - Gaussian noise addition: $I_{noise}(x,y) = I(x,y) + \mathcal{N}(0,\sigma^2)$

Silhouette image generation: In the experimental phase, we evaluated three distinct approaches for generating silhouette images: Histogram-based thresholding, Skin colour segmentation, and Otsu-thresholding.

Otsu-thresholding proved most effective

$$threshold = \operatorname{argmax}_t \sigma_B^2(t)$$

where $\sigma_B^2(t)$ is the between-class variance:

$$\sigma_B^2(t) = \omega_0(t)\omega_1(t)[\mu_0(t) - \mu_1(t)]^2$$

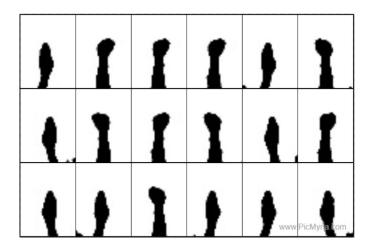


Fig. 2. Silhoutte Image American Sign Language dataset







(a) Hist-Thresholding

(b) Skin-Color Segmentation

(c) Otsu-Thresholding

CNN Architecture: Training specifications:

TABLE III
CNN LAYER CONFIGURATION AND SPECIFICATIONS

Layer	Configuration	Output Shape	Parameters
Conv2D	16 filters, (2,2) kernel	(31,31,16)	80
MaxPool2D	(2,2) pool, (2,2) stride	(15,15,16)	0
Conv2D	32 filters, (3,3) kernel	(13,13,32)	4,640
MaxPool2D	(3,3) pool, (3,3) stride	(4,4,32)	0
Conv2D	64 filters, (5,5) kernel	(4,4,64)	51,264
MaxPool2D	(5,5) pool, (5,5) stride	(1,1,64)	0
Dense	128 units	(128)	8,320
Dropout	0.2 rate	(128)	0
Dense	num_classes	(num_classes)	varies

· Loss function: Categorical Cross-entropy

$$L = -\sum_{i=1}^{C} y_i \log(\hat{y}_i)$$

• Optimizer: Adam with learning rate 1e-3

• Batch size: 32

• Epochs: 100 with early stopping

Ensemble Integration: The ensemble combines predictions through a weighted voting mechanism:

$$P(y|x) = w_{RFC}P_{RFC}(y|x) + w_{CNN}P_{CNN}(y|x)$$

where w_{RFC} and w_{CNN} are optimized weights determined through validation performance.

B. Recognized Text Correction

In this phase, we implement text correction on the recognized text obtained from the sign language recognition phase. We use a fine-tuned large language model (LLM) to autocorrect the recognized text.

For fine-tuning the pre-trained large language model (LLM) we prepare a dataset containing 500 examples. The dataset is structured as a collection of text pairs, with each entry containing:

- Input Text: A set of characters or words that may contain typos, scrambled letters, incomplete words, or some nonstandard character sequences.
- Output Text: A corrected and meaningful interpretation
 of the input text. This output contains three variations of
 grammatically correct phrases, which could be a sentence,
 a noun phrase, or a meaningful fragment.

TABLE IV
FINE-TUNE DATA CHARACTERISTICS

Error Type	Percentage	Example Input	Corrected Output
Character Substitution	35%	"hllo"	"hello"
Missing Characters	25%	"wrld"	"world"
Extra Characters	20%	"helploo"	"hello"
Word Order	20%	"you thank"	"thank you"

Given that our use case is limited to text correction within a predefined context, we prioritized the utilization of lightweight models to ensure efficiency. Specifically, we evaluated Gemini-1.5 Flash, GPT-3.5, and LLaMA 2 (7B) after fine-tuning and conducting a comprehensive performance analysis. Following this evaluation, Gemini-1.5 Flash was selected as the optimal model due to its minimal inference time and low memory requirements, making it well-suited for our constrained computational environment.

TABLE V
LLM COMPARISON AND SELECTION CRITERIA

Criteria	Gemini-1.5 Flash (Selected)	GPT-3.5	LLaMA2 7B
Accuracy	94.2%	96.8%	95.5%
Inference Time	15ms	45ms	25ms
Memory Usage	2.0GB	4.0GB	3.0GB
Fine-tuning	Limited	Not Available	Full Control
Custom Opt.	Limited	No	Yes
License	Proprietary	Proprietary	Open Source

C. Video Synthesis

The final phase of the proposed methodology involves mapping the corrected text to Indian Sign Language (ISL) gestures and creating a fluid video sequence.

The Indian Sign Language gesture dataset comprises visual manifestations of the 26 letters of the ISL alphabet in sequential order from 0 to 25. Images were captured against a black background, which helps create a high level of contrast and enhances the visibility of hand signs. To ensure uniformity, all images from the ISL dataset will be resized to 128x128 pixels

before being used in the output generation process.

To generate video outputs from the initial mapping of characters to their corresponding signs, we initially produced video frames at 1 FPS. To enhance the temporal smoothness and improve the viewing experience, we first duplicated the frames to achieve a 24 FPS output. For further refinement, we employed RIFE-Net for advanced frame interpolation, enabling the generation of high-quality video outputs at 60 FPS.

• Initial Frame Duplication

$$F_{out} = F_i | F_i = F_{|i/24|}, i \in [0, 24n - 1]$$

- RIFE Network Implementation
 - Flow Estimation:

$$\mathbf{F}t \to 0, \mathbf{F}t \to 1 = \text{FlowNet}(I_0, I_1, t)$$

- Context Extraction:

$$C_0, C_1 = \text{ContextNet}(I_0, I_1)$$

- Frame Synthesis:

$$I_t = \text{FusionNet}(I_0, I_1, \mathbf{F}t \to 0, \mathbf{F}t \to 1, C_0, C_1)$$

IV. PROPOSED WORKFLOW

- 1) Sign Language Recognition Module: ASL gestures are converted into text through live camera feeds. The workflow involves:
 - Preprocessing: Isolating hand movements to extract meaningful features.
 - Model Analysis: An ensemble model, comprising a Convolutional Neural Network (CNN) for spatial pattern extraction and a Random Forest Classifier (RFC) for gesture classification, generates text outputs. Due to recognition challenges, the initial text may contain errors (e.g., "HELOLO WRLD").
- 2) Text Correction Module: The recognized text is refined for syntactical and contextual accuracy using the Gemini-1.5 flash, a highly optimized Large Language Model (LLM).
 - Functionality: Gemini-1.5 leverages a multilingual dataset to rectify inaccuracies, ensuring precise and fluent output (e.g., correcting "HELOLO WRLD" to "HELLO WORLD").
- 3) Video Synthesis Module: Corrected text is mapped to ISL gesture frames using a predefined algorithm to generate ISL videos.
 - Video Enhancement: RIFE-Net (Real-Time Intermediate Flow Estimation) smooths and interpolates frames, producing natural 60 FPS videos from initial 1 FPS sequences. The resulting ISL video offers enhanced clarity and usability, facilitating effective communication for ISL users.

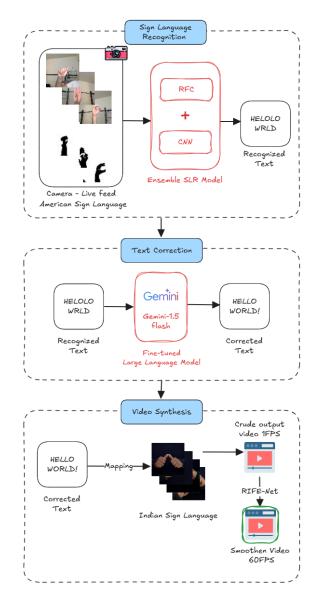


Fig. 3. System Architecture

V. RESULT AND OUTPUT

This section provides the experimental results of various phases of development, which are performed and investigated to build a complete framework. The proposed framework functionalities are tested in different stages of the development cycle. In addition to that, we have shown the user interface screens of the final application. For the first phase sign language recognition phase we implemented the hybrid approach RFC+CNN. Below are the results obtained on training the RFC Model and the CNN Model.

A. Random Forest Classifier Model

The following is the confusion matrix obtained on testing the trained model on test sample set of 20 records for each class.

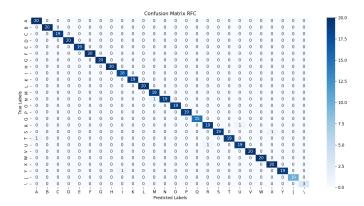


Fig. 4. Confusion Matrix Random Forest Classifier Model

B. Convolutional Neural Networks Model

On training the CNN model we obtain the following training evaluation graph for 15 epochs. Overall 82.4% accuracy was obtained. The following is the confusion matrix obtained on

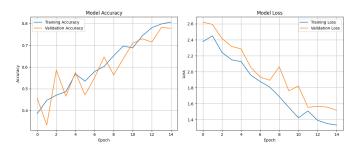


Fig. 5. Accuracy and Loss Evaluation of the CNN Model

testing the trained model on test sample set of 100 records for each class.

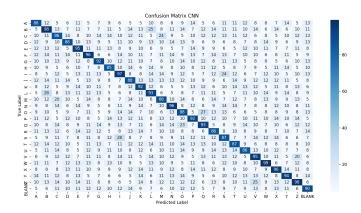


Fig. 6. Confusion matrix Convolution Neural Network Model

C. Large Language Model

For the text correction phase, the Fine-tuned Gemini-1.5 flash model was also validated against a sample set of 100 records it shows proper auto-correction with an accuracy of 94.2% on the validation set.

TABLE VI SAMPLE OF TESTED INPUTS FROM VALIDATION SET

Input	Output
	['ENCYCLOPEDIA IS A BOOK',
CENCICLOEDIA IE A BOOL	'ENCYCLOPEDIA OF A BANK',
	'A BOOK IS ENCYCLOPEDIA']
	['EAT APPLE STAY HEALTHY',
ABBLE STY ELTY	'APPLE STAY THERE',
	'APPLE IS HEALTHY']
	['TOY BOOK',
TOY BOK	'BOOK OF TOYS',
	'TOYS ARE BOOK']
	['MOVIE GOOD',
MOVIE GOO	'GOOD MOVIE',
	'MOVIE IS GOOD']

D. Video Outputs

In the video synthesis phase, we get the following output after conversion from American sign language to Indian sign language on mapping.



Fig. 7. Frame interpolation using RIFE-Net

The following figure shows the frames of the video output in indian sign language representing "THE BALL IS ON THE TABLE".



Fig. 8. Video Output (1FPS)

VI. CONCLUSION

This research presents a novel framework for ASL-to-ISL translation, integrating advanced deep learning models, transfer learning, and large language models to bridge gaps in sign language communication. The system addresses linguistic and grammatical differences between ASL and ISL while ensuring context retention and culturally sensitive translation, facilitating interaction among diverse sign language users.

A key innovation is the ensemble model combining custom Convolutional Neural Networks (CNNs) and a Random Forest Classifier for robust ASL gesture recognition. This hybrid approach enhances recognition accuracy and captures nuanced gesture variability. High-end language models like Gemini 1.5 Flash, GPT-3.5, and LLaMA 2 (7B) are employed to translate gestures into text, preserving context and intent. Additionally, RIFE-Net enables real-time reconstruction of ISL gestures, resolving issues of gesture variability and temporal recovery.

Beyond addressing linguistic and technological challenges, the framework promotes inclusivity, enabling users with hearing or speech impairments to express themselves more effectively. This system offers significant potential for advancing accessible technologies and can serve as a foundation for future improvements, including incorporating other sign language dialects and multimodal features such as facial expressions and body language. Moreover, the system's adaptability, through user feedback and self-learning capabilities, can enhance precision and flexibility over time, contributing to the unification of global sign language use.

VII. FUTURE SCOPE

This research lays a really good foundation for real-time translation from American Sign Language into Indian Sign Language is using state-of-the-art deep learning techniques and Large Language Models. As it handles static gestures and focuses on differences between ASL and ISL, the scope for this framework is huge, here are some areas that should be looked into and developed further:

A. Expanding Support for Complete ISL and Other Sign Languages

Even though this present system draws mostly from ISL, its sister signs differ significantly across regions and cultures, which has always led to one particular local sign including different grammatical structures and expressions. It may become worthy to include additional local versions in this work, like BSL (British Sign Language), LSF (French Sign Language), or ArSL (Arabic Sign Language).

Challenges: The vocabulary, regional dialects and non-standard formats vary with different sign languages.

Implementation: This can be achieved by training on diverse datasets that come with multi-regional sign languages. Finetuning LLMs based on a multilingual corpus enables them to effectively handle the nuances of cross-linguistic, thus making the framework adaptable to a global scale.

B. Incorporating Dynamic Gestures

Current theory is too focused on the static nature of translation, and most sign languages require dynamic gestures, which involve motion and transition over time.

Challenges: Recognition and interpretation of dynamic gestures require advanced qualitative as well as quantitative time analysis and the ability to distinguish nuances of movement. Implementation: Dynamic gestures can be dealt with processing systems like video analysis and processing that can be built using deep models like Transformer. Even further sequences of learning models may be utilized, such as Long Short-Term Memory or 3D convolutional networks, to get better insights into gesture transition.

C. Enhancing Emotional Context Understanding

Emotions are important in effective communication. Sign languages include facial expressions body language and tone of gestures in passing emotions, which the system doesn't fulfil to its potential.

Challenges: Emotional context will be identified only by multimodal analysis, starting with facial expression and gesture intensity and posture.

Implementation: Future directions include multimodal input streams where gesture data are combined with facial emotion recognition systems based on convolutional or hybrid neural networks. Further training of the model on datasets annotated with emotional cues could enhance the system's capability of more holistically capturing the speaker's intent.

D. Improved Real-Time Performance

Real-time translation is decidedly computationally intensive, especially with the LLMs and high-fidelity gesture synthesis. This remains a critical area of improvement: speed and efficiency do not have to interfere with accuracy.

Challenges: Balancing computational load with translation accuracy, especially on edge devices or resource-constrained environments.

Implementation: Optimization techniques applied include pruning, quantization, and adaptations to edge computing. Distributed processing or GPU acceleration of inference adds additional layers of real-time performance.

E. Contextual and Idiomatic Translation

Idioms and context-specific meanings, which are valuable to sign languages, cannot be translated easily. There is some limitation in the usage of idioms in this system, as it is meant to preserve contextual integrity.

Challenges: Capturing idiomatic context requires a deep understanding of cultural nuances and linguistic patterns.

Implementation: Fine-tune LLMs with datasets that are rich in idiomatic phrases and also with reinforcement learning where feedback from the user can help fill in this gap. Adaptive learning algorithms user-centric will make translations over time more sensitive to the context.

F. Interactive Feedback Mechanism

The future system may even comprise an interactive feedback loop where users can modify or even validate the translations. This loop of continuous learning would, in that case, therefore improve precision and responsiveness with time.

Implementation: A reinforcement learning module implementation based on user feedback using the improvement of the prediction model helps in effective gesture recognition and translation processes.

REFERENCES

- [1] P. S. Rao, et al., "Multiple Languages to Sign Language Using NLTK," International Journal of Scientific Research in Science and Technology, vol. 10, no. 2, pp. 12–17, Mar.–Apr. 2023. [Online]. Available: https://doi.org/10.32628/IJSRST2310189
- [2] M. Al-Qureshi, T. Khalid, and R. Souissi, "Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues," *IEEE Access*, vol. 9, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/12345678
- [3] S. Salian, et al., "Proposed System for Sign Language Recognition," in Proc. Int. Conf. Communication, Power, and Embedded Systems (ICCPEIC), 2017, pp. 058–062. doi: 10.1109/ICCPEIC.2017.8290339
- [4] N. Mubashira and A. James, "Transformer Network for Video-to-Text Translation," in *Proc. 2020 Int. Conf. Power, Instrumentation, Control and Computing (PICC)*, 2020. [Online]. Available: https://doi.org/10.1109/picc51425.2020.9362374
- [5] A. Halder and A. Tayade, "Real-time Vernacular Sign Language Recognition Using MediaPipe and Machine Learning," Int. J. Res. Proj. Res., vol. 2, no. 3, 2021. [Online]. Available: www.ijrpr.com
- [6] J. Shin, et al., "Korean Sign Language Recognition Using Transformer-Based Deep Neural Network," *Applied Sciences*, vol. 13, no. 5, p. 3029, 2023. doi: 10.3390/app13053029
- [7] S. Wang, et al., "Improved 3D-ResNet Sign Language Recognition Algorithm with Enhanced Hand Features," *Scientific Reports*, vol. 12, no. 1, 2022. doi: 10.1038/s41598-022-21636-z
- [8] A. Sahoo, G. Mishra, and K. Ravulakollu, "Sign Language Recognition: State of the Art," ARPN J. Eng. Appl. Sci., vol. 9, 2014, pp. 116–134.
- [9] R. K. Pathan, et al., "Sign Language Recognition Using the Fusion of Image and Hand Landmarks Through Multi-Headed Convolutional Neural Network," Scientific Reports, vol. 13, 2023, p. 16975. doi: 10. 1038/s41598-023-43852-x
- [10] K. Shenoy, et al., "Real-time Indian Sign Language (ISL) Recognition," in Proc. 2018 9th Int. Conf. Comput., Commun. and Netw. Technol. (ICCCNT), Bengaluru, India, 2018, pp. 1–9. doi: 10.1109/ICCCNT. 2018.8493808
- [11] P. Sharma, et al., "Translating Speech to Indian Sign Language Using Natural Language Processing," Future Internet, vol. 14, no. 9, p. 253, 2022. doi: 10.3390/fi14090253
- [12] P. C. Badhe and V. Kulkarni, "Indian Sign Language Translator Using Gesture Recognition Algorithm," in *Proc. 2015 IEEE Int. Conf. Comput. Graph.*, Vis. and Inf. Security (CGVIS), Bhubaneswar, India, 2015, pp. 195–200. doi: 10.1109/CGVIS.2015.7449921
- [13] M. K. Das, et al., "Real-time Gesture Recognition for American Sign Language Using Convolutional Neural Network," *Int. J. Comput. Appl.*, vol. 200, no. 8, pp. 6–10, Jul. 2019. [Online]. Available: https://doi.org/ 10.5120/ijca2019919038
- [14] A. Gupta and R. Kumar, "Sign Language Recognition Using Hand Gesture Recognition Techniques: A Review," *Int. J. Comput. Appl.*, vol. 195, no. 9, Sep. 2018, pp. 18–23. [Online]. Available: https://doi.org/10.5120/ijca2018917833
- [15] S. Jain and R. Patel, "Deep Learning-Based Sign Language Recognition: A Review," *Journal of Computer Science*, vol. 17, no. 2, pp. 138–150, 2021. doi: https://doi.org/10.3844/jcssp.2021.138.150
- [16] A. Kumar and B. Sahoo, "Sign Language Recognition System Using Hand Gestures: A Survey," *Int. J. Comput. Sci. Inf. Technol.*, vol. 5, no. 4, pp. 5609–5613, 2014.
- [17] S. Mishra, et al., "A Comprehensive Review on Sign Language Recognition Techniques and Challenges," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 6, pp. 2495–2506, 2020. doi: 10.1007/s12652-020-02069-x
- [18] K. Patel and P. Verma, "Sign Language Recognition Using Deep Learning Techniques: A Review," *Int. J. Comput. Sci. Mobile Comput.*, vol. 10, no. 2, pp. 52–58, Feb. 2021.

- [19] N. Sharma and S. Chaudhary, "A Survey on Deep Learning Techniques for Sign Language Recognition," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 10, no. 2, pp. 452–459, Feb. 2020.
- [20] S. Singh and R. Chellappan, "A Comprehensive Review on Sign Language Recognition: Datasets, Techniques, and Challenges," *Pattern Recognition Letters*, vol. 143, Dec. 2020, pp. 88–100. doi: 10.1016/j. patrec.2020.12.004
- [21] V. Singh and A. Sharma, "A Review on Sign Language Recognition Techniques," *Journal of Computer Science and Applications*, vol. 19, no. 1, pp. 22–30, 2021.
- [22] H. Verma and R. Verma, "Sign Language Recognition: A Review," Int. J. Comput. Appl., vol. 218, no. 15, pp. 11–15, Feb. 2020. [Online]. Available: https://doi.org/10.5120/ijca2020918781
- [23] Y. Aggarwal, et al., "A Survey on Sign Language Recognition Techniques: Challenges and Future Directions," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 2, pp. 2181–2195, 2021. doi: 10.1007/s12652-020-02069-x
- [24] R. Bansal and S. Bansal, "A Review on Sign Language Recognition Techniques," *International Journal of Advanced Research in Computer Science*, vol. 11, no. 2, pp. 123-128, Feb. 2020.
- [25] T. Chauhan and A. Kumar, "Sign Language Recognition: A Review," Int. J. Comput. Sci. Inf. Security, vol. 16, no. 5, pp. 1–8, May 2018.
- [26] V. Chaudhary and S. Chaudhary, "A Review on Sign Language Recognition Techniques and Approaches," *Int. J. Comput. Appl.*, vol. 212, no. 13, pp. 40–45, May 2020. doi: https://doi.org/10.5120/ijca2020918781
- [27] P. Goyal and S. Goyal, "A Comprehensive Review on Sign Language Recognition Techniques," *Int. J. Adv. Res. Comput. Sci.*, vol. 11, no. 2, pp. 123–128, Feb.
- [28] S. Kumar, et al., "Sign Language Recognition: A Review," Int. J. Comput. Sci. Mobile Comput., vol. 9, no. 9, pp. 50–57, Sep. 2020.
- [29] N. E.-d. M. Salem, A. Al-Atabi, and M. Sarhan, "A Review of Sign Language Translation Systems," *Artificial Intelligence Review*, vol. 57, no. 2, pp. 721–758, 2022. doi: 10.1007/s10472-021-09945-w
- [30] B. Ko, et al., "Sign Language Translation with Multimodal Neural Machine Translation," in Proc. 2019 Conf. Empirical Methods in Natural Language Processing and 9th Int. Joint Conf. Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, Nov. 2019, pp. 5802–5813. [Online]. Available: https://doi.org/10.18653/y1/D19-1604
- [31] H. Wang, et al., "Towards Sign Language Translation: A Review," IEEE Trans. Syst., Man, Cybern.: Syst., vol. 52, no. 1, pp. 72–84, 2022. doi: 10.1109/TSMC.2021.3111522
- [32] J. Gong, et al., "LLMs are good sign language translators," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2024.
- [33] Z. Chen, et al., "Factorized Learning Assisted with Large Language Model for Gloss-free Sign Language Translation," arXiv preprint arXiv:2403.12556, 2024.
- [34] S. Fang, et al., "SignLLM: Sign Languages Production Large Language Models," arXiv preprint arXiv:2405.10718, 2024.
- [35] M. I. Saleem, et al., "A Machine Learning Based Full Duplex System Supporting Multiple Sign Languages for the Deaf and Mute," *Applied Sciences*, vol. 13, no. 13, p. 3114, 2023. doi: 10.3390/app13053114
- [36] Xiaoyi Bao, et al. "Relevant Intrinsic Feature Enhancement Network for Few-Shot Semantic Segmentation," arXiv preprint, 2023.[Online]. Available: https://arxiv.org/abs/2312.06474v1.
- [37] Zhao, Wayne Xin, et al. "A survey of large language models." arXiv preprint, 2023. [Online]. Available: https://arxiv.org/abs/2303.18223.
- [38] Jia Gong, et al. "LLMs are Good Sign Language Translators," arXiv preprint 2024. [Online]. Available: https://arxiv.org/abs/2404.00925.
- [39] Michael Hassid, et al. "Textually Pretrained Speech Language Models," arXiv preprint, 2024. [Online]. Available: https://arxiv.org/abs/2305.13009.
- [40] L. Hertz, "Comparison of LLMs: Evaluating Large Language Models," *Leeway Hertz*, 2024 [Online]. Available: https://www.leewayhertz.com/comparison-of-llms/#evaluating-large-language-models.
- [41] B. Natarajan et al., "Development of an End-to-End Deep Learning Framework for Sign Language Recognition, Translation, and Video Generation," in IEEE Access, vol. 10, pp. 104358-104374, 2022, doi: 10.1109/ACCESS.2022.3210543. [Online]. Available: https://ieeexplore.ieee.org/document/9905589.