NRPU Proposal.txt

by saleem khan

Submission date: 19-Oct-2025 01:18AM (UTC-0700)

Submission ID: 2753563543

File name: NRPU_Proposal.txt (18.6K)

Word count: 2579

Character count: 16332

Problem Statement (Max. 500 words)

Existing PSL datasets are often limited in scope, focusing primarily on isolated signs rather than continuous sign sequences, which hinders the development of robust translation models. Current transformer-based SLT models show promise but need more data. A significant challenge is bridging the semantic gap between text encoding, key-point encoding, and video encoding. Transformer-based SLT models require high computational resources, making real-time deployment challenging. Furthermore, data set limitations pose a significant bottleneck. Many sign language datasets are small, unbalanced, and lack diversity, leading to models that are biased toward specific signers and fail to generalise well across different environments. The main research gap lies in finding better ways to develop optimised multi-modal fusion strategies, improve computational efficiency through lightweight architectures, and expand dataset diversity to build more generalisation SLT models. These results will help create highly accurate and scalable sign language translation systems that make sign language communication more accessible to a wider audience. Publicly available large-scale PSL datasets are scarce. Need for models handling continuous sign-to-Urdu translation

*

Objectives and Scope (Max. 500 words)

Sign language is a sophisticated visual language system that conveys meaning using body language, facial expressions, and hand gestures. Sign language is the main mode of communication for millions of people around the world who are deaf or have hearing impairments. According to the World Federation of the Deaf, Pakistan has estimated 10.2 million deaf people, over 70 million people worldwide use various sign languages. Existing PSL datasets are often limited in scope, focusing primarily on isolated signs rather than continuous-sign sequences, which hinders the development of robust translation models. Using only RGB features to analyze expressive actions such as face expressions, lips and eye motions, head movements, and body postures, along with physical actions like hand gestures and arm positioning, can be restrictive because different datasets have different backgrounds and signers. Despite this drawback, the RGB characteristics are the only ones used in most Sign Language Translation (SLT) studies. In order to properly capture the posture and arrangement of the body parts engaged in sign language gestures, we employed key point features in addition to RGB features for better sign language translation as it also captures the spatio-temporal features. Transformers, which are effective in focusing on the most pertinent video frames and capturing a broader, higher-level context, have also been used in most SLT research projects. However, the natural graph structure of sign language is overlooked and does not capture minute nuances. Limited sign language datasets lead to overfitting and poor generalization.

The objectives of this research are to fill the gap by creating a continuous Pakistan Urdu sign language data set and the use of advanced deep learning techniques to enhance the continuous translation of sign language. To solve this, we used a combined encoding method that captured the temporal and spatial dependencies on skeleton graphs as well as the context of sign language expressions utilizing a UrduBert transformer and STGCN architecture for primary motion modeling, fusing with transformer encoders for improved translation accuracy, and expanding dataset diversity. Our approach, Hybrid-STGCN, demonstrates its efficacy in improving translation accuracy across various sign languages by experimentally achieving competitive performance in the Pakistan Sign Language dataset. In summary ,the objective of this research is to address the challenges faced by the low resource Urdu language:

- Fill the gap by creating a continuous Pakistan Urdu sign language dataset.
- Enhance continuous translation in sign language while addressing key challenges related to multi-modal feature integration, computational efficiency, and fusion strategies.
- To improve translation accuracy by using the UrduBERT transformer for improved contextual and semantic representation in Pakistan Sign Language (PSL) translation

*

Research Significance and the Challenges Addressed (Max. 500 words)

This project targets an urgent national and global accessibility need: seamless communication between Deaf/hard-of-hearing signers and non-signers across daily life like education, healthcare and emergencies, workplaces, public services, online platforms, and smart environments. By enabling real-time translation of Pakistan Sign Language (PSL) into text/speech, the work advances inclusive digital transformation, supports SDG-3 (Good Health & Well-Being), SDG-4 (Quality Education), SDG-8 (Decent Work), SDG-9 (Industry, Innovation & Infrastructure), SDG-10 (Reduced Inequalities), and directly contributes to Pakistan's accessibility and e-governance priorities. Practically, the same core technology can power classroom tools, hospital triage and telehealth, customer support, public counters (airports, courts, tax offices), video conferencing, social media captioning, and smart-home/IoT command interfaces, as well as media, gaming, and tourism way-finding.

Scientifically, the project addresses a long-standing gap in continuous PSL translation. Most prior efforts focus on isolated signs and RGB-only cues, limiting robustness to signer, background, and lighting variations. Our approach advances the field along three dimensions:

1. Data contribution (localization & scale): We will build a continuous PSL corpus capturing realistic conversational flow and regionally specific lexicon. This fills a critical low-resource gap, enabling research and local startups to innovate on top of an ethically collected, diverse dataset.
2. Model innovation (multimodal, structured, and context-aware): We propose Hybrid-STGCN to jointly encode skeletal/keypoint graphs (for fine-grained hand, face, and body articulation) and RGB features (for appearance cues), fused with transformer encoders for long-range temporal dependencies and UrduBERT for linguistically faithful target generation. This unites graph-based motion modeling with sequence-to-sequence language understanding which is crucial for continuous, co-articulated signing.
3. Deployment realism (from lab to field): We emphasize real-time inference, signer-independence, domain adaptation, and edge-friendly optimization to enable on-device or near-device use in clinics, classrooms, kiosks, and mobile apps.

Key Challenges Addressed:

1. Data scarcity & diversity: Lack of continuous PSL corpora, regional variation, and coarticulation. We mitigate via targeted data collection,

balanced signer demographics, varied settings, and ethical annotation protocols.

2. Multimodal fusion & temporal reasoning: Properly combining RGB with skeleton graphs and modeling long-range dependencies. Our Hybrid-STGCN + transformer fusion explicitly captures spatio-temporal structure and discourse context.
3. Generalization & robustness: Handling different signers, speeds, lighting, occlusions, and backgrounds. We adopt augmentation, domain-adversarial training, and calibration for robust performance.
4. Real-time constraints: Achieving low latency for live interpretation in hospitals, emergencies, and service counters. We apply model compression, quantization, and efficient backbones suited for mobile/edge hardware.
5. Language generation fidelity: Producing grammatically and semantically accurate Urdu text/speech from visual input. UrduBERT fine-tuning with alignment strategies and curated parallel references will reduce hallucinations and preserve meaning.

Evaluation & ethics: Beyond BLEU/WER, we will conduct user-centered evaluations with Deaf communities and practitioners, ensure informed consent, privacy/security of recordings, and bias audits.

Expected Significance:

Deliverables include a publicly shareable continuous PSL dataset (with consent), state-of-the-art multimodal translation models, and pilot deployments (clinic/classroom prototypes). The outcome is a scalable accessibility platform that can be integrated into e-services, social media apps, EdTech, HealthTech, and Assistive AI products, positioning Pakistan as a regional leader in inclusive, AI-powered communication technologies.

*

Research Methodology (Max. 1000 words)

We present a detailed explanation of our proposed method for the translation of sign language. The approach is designed to effectively bridge the gap between visual gestures and meaningful text output by leveraging advanced machine learning and deep learning techniques. We outline the architecture, key components, data processing stages, and the models utilised to ensure accurate translation from sign language videos to the corresponding textual descriptions. The following subsections will discuss each part of the proposed methodology, highlighting the innovation, design choices, and performance considerations that contribute to the effectiveness of our system.

1. Dataset Preparation

The research seeks to construct a detailed data set for Pakistan Sign Language (PSL) by collecting, processing and structuring data from video sources. The data set will help develop models that translate continuous sign languages by providing diverse representations between different signers, environmental differences, and various sign styles. The approach consists of three main steps: data collection, pre-processing, and data set structure.

1.2 Data Collection

A large dataset was collected from the Pakistan Sign Language (PSL) site, which serves as a detailed PSL dictionary. The dataset contains 79 folders with an overall total of 7797 videos,with different PSL signs shown in each video. Each video is provided with correspondence in English and Urdu text annotations, providing a bilingual dataset for training and evaluation. Use 1,000 Basic Signs in 7 languages: English, Urdu, Punjabi, Sindhi, Pashto, Balochi, and PSL.
Contents include: 6,000 Word PSL Dictionary, PSL Animated Stories, Tutorials, Student Tutorials, and more. This data set is considered crucial in building an efficient PSL translation model; it has provided a diverse set of signs for training and evaluation that include different hand movements, facial expressions, and body postures. The availability of both English and Urdu text annotations also enhances the functionalities of the multilingual translation of this dataset, allowing it to be used for other domains such as sign recognition and natural language processing. The well-structured organization of this dataset eases the pre-processing, feature extraction, and model training steps, leading to future directions in PSL translation and accessibility technologies. Sign languages encompass isolated signs (word-level) and continuous PSL sentences (sentence-level translation).

1.2 Data Pre-processing

The source video data of sign language are usually noisy and inconsistent themselves, they may come with background noises, superfluous frames, non-uniform signing speeds and variations in hand gestures. Therefore, pre-processing is very important for such data to ensure a certain level of standardization for the whole dataset. The uses these pre-processed data to create cleaned, structured data. These pre-processing steps include the following.

1. Video Frame Extraction: Convert video into sequences of frames (for example, 30 FPS or adaptively depending on gesture speed).
2. Frame resizing and normalization: frames in a fixed dimension (e.g., 224×224 pixels) to ensure a uniform input size. Normalization of pixel values for better neural network performance.
3. Hand/Body Pose Estimation: Use MediaPipe/OpenPose to extract key-points.
4. Data Augmentation: Flip, rotate, or slightly shift frames to generalise the model. The pre-processing steps thus improve the quality of the data set in terms of consistency and model performance in PSL tasks. After pre-processing, the dataset is organized into structured categories:

a. Word-Level PSL Videos (Single-word signs with gloss labels).
b. Sentence-Level PSL Videos (Full PSL sentences with translated text).
c. Keypoint-Based Representations (Skeleton-Based Sign Movement Tracking).
d. Metadata (Signer demographics, environment conditions, and video details).

To ensure balance, the data set is adjusted to maintain an equal distribution of common PSL words and sentences, preventing the model bias towards frequently occurring signs.

2. Feature Extraction

Feature extraction in Pakistan Sign Language (PSL) translation involves capturing key visual and motion-based information from sign language videos to improve recognition accuracy. To capture the motion and structure of sign gestures, we use a two-stream deep learning approach:
 Spatio-Temporal Feature Extraction using I3D: The Inflated 3D Convolutional Network (I3D) is used to extract spatio-temporal features from the RGB frames.

Keypoint Feature Extraction using ST-GCN: To further improve accuracy, we extracted human skeletal key-points using OpenPose or MediaPipe. These key-points are then processed using a spatial temporal graph convolutional network (ST-GCN) to effectively model body movements and hand gestures.

Feature Fusion: The I3D-extracted features and ST-GCN key-point representations are fused into a single feature vector, providing a richer representation of the sign language gestures. In this approach, we deal with the multi-modal approach for continuous sign language translation. Thus, to increase the accuracy of translations, RGB video features, and key-point-based skeletal features are embedded into a single training framework using two-stream fusion. The fused representation will then pass through the Transformer decoder with Urdu BERT Embeddings, which accepts text embeddings and outputs translated spoken language text from the sign input. The final output will incorporate a post-processing mechanism aiming to achieve fluency and grammatical correctness, improving robustness for real-world scenarios. The process of this methodology accurately connects the semantic gap between video, key-point, and text encoding, thus improving the more accurate and efficient sign language translation.

3. Training the models

To develop high-performing models to classify sign and text, we follow deep learning models for continuous sign language translation (SLT): Input Pre-processed Video Frames: The RGB video frames are processed to extract both RGB and key-point features for sign language representation. Feature extraction using I3D and STGCN: I3D extracts motion features from video frames, STGCN-LSTM captures spatial-temporal dependencies in keypoint-based skeletal graphs. Feature Fusion: Concatenate I3D and ST-GCN feature representations into a single sequence representation vector. Transformer Encoder: Takes the fused feature vector as input. Transformer Decoder with Urdu BERT Embeddings: The decoder generates Urdu words, one at a time. Text Post-processing: De-tokenize the generated tokens. Apply true casing (restore the correct case for words).
 Output: Generated human-readable Urdu sentence.

4. Evaluation of models

BLEU Score and Word Error Rate (WER) will be used to evaluate the model accuracy. The Bilingual Evaluation Understudy (BLEU) score is utilized in translation
models to evaluate the quality of predicted sign sequences by comparing them with reference translations.

*

Expected Outcomes (Max. 500 words)

The expected outcome of our project is that we will develop and deploy an AI-powered system aimed at narrowing the gap between the deaf, mute, or both, and the hearing world. We will deliver a complete mobile app-based on our research for continuous Pakistan Sign Language (PSL) translation. First, we will curate and release a continuous PSL dataset comprising ethically collected videos with synchronized RGB and skeletal keypoints, Urdu transcriptions, and rich metadata, which will enable reproducibility and responsible reuse. Building on this foundation, we will develop Hybrid-STGCN–UrduBERT translation models that support PSL Urdu text/speech and the reverse pathway (Urdu to PSL avatar/gesture plan). The inference stack will be optimized for real-time use, with quantized variants engineered for edge devices, and exposed via production-grade APIs/SDKs to simplify Android and Web integration. We will develop reference applications (mobile and web) that provide live camera input, subtitles, Urdu text-to-speech, session logs, and accessibility features such as adjustable caption sizes and configurable signing rate.

We will evaluate the system rigorously. Quantitatively, we aim to improve continuous-translation scores over RGB-only baselines. Robustness will be tested for signer, background, lighting, and occlusion shifts, targeting ≤10% WER degradation and strong retention under stress tests, with cross-domain generalization capped at ≤15% drop from in-domain performance. Equally important, we will conduct user-centered studies with Deaf community partners, educators, speech therapists, and frontline staff, aiming for a System Usability in real scenarios (e.g., classroom Q&A, clinic triage, public service counters). All releases will include ethics and safety checklists, bias audits, and detailed model/data.

To drive adoption, we will run at least three sector pilot test such as in education (lecture/classroom captioning), healthcare , and a public-service, with SOPs, staff training, and feedback loops for continuous model updates. Integration pathways will include WebRTC-based video conferencing plugins, connectors for customer-support platforms, and a roadmap for smart kiosks and IoT triggers. We will share a pre-trained checkpoints, core code, and the in-house annotation tooling to catalyze local research and innovation. The team will target one to two peer-reviewed publications in leading CV/NLP domains. Alongside, we plan to mentor 10–15 student researchers and run workshops that build national capacity in accessible, responsible AI.

In the longer term, the project establishes a scalable accessibility layer for Urdu/PSL with clear extensions to regional languages, gesture-to-gesture communication without text intermediaries, wearable integrations, and fully on-device operation. Its modular architecture lowers the cost of expanding to new domains and populations, while measurable accuracy gains and real-time performance ensure practical value. By aligning with national digital inclusion priorities and multiple SDGs, the work provides evidence to inform accessibility standards for public-facing services in Pakistan and positions local institutions to lead in AI-powered assistive communication.

*

Similarity Index along with Additional Details (if any)

*

NRPU Proposal.txt

ORIGINALITY REPORT

2%	2%	1%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.evsu.edu.ph Internet Source	1%
2	psl.org.pk Internet Source	1%
3	deepai.org Internet Source	<1%
4	www.coursehero.com Internet Source	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		