UNIVERSITY OF TORONTO

American Sign Language Translation

Team 78: Amelie Wang; Angela Yang; Bill Qu; Yiming Zhang; Supervisor: Ben Liang; Administrator: John Taglione

Project URL:

Pdf Poster:

Background and Motivation

- **What is Sign Language? :** Sign language is a visual way of communicating using hand gestures, facial expressions, and body language instead of spoken words.It is mainly used by people who are deaf or hard of hearing.
- **Motivation**: To bridge the communication gap for the deaf and hard-of-hearing community by enabling seamless, real-time sign language translation in digital interactions

Project Requirements

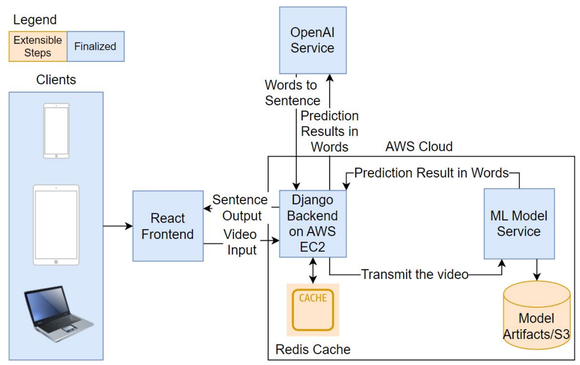
- **Accuracy**: At least **60%** accuracy in recognizing and translating **100** gestures, even in challenging conditions.
- **User-Friendliness**: Simple, intuitive interface with no training required. On-screen prompts and visual feedback improve usability, verified through user testing.
- **Speed**: Translations within **5** seconds via an efficient end-to-end pipeline. No current product offers instant live translation without adjustments.
- **Reliability**: Consistent performance for **3+** hours, tested with **40** users in varied conditions (incomplete gestures, background movement, similar gestures, different angles).
- **Integration**: Compatible with **Windows, macOS, Linux, Android,** and **iOS** across **smartphones, tablets**, and **computers** for everyday use.

Technical Innovations

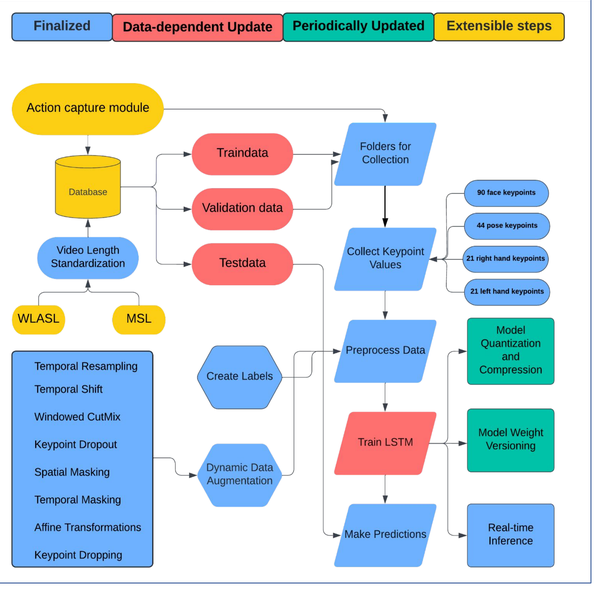
- 8-head attention simultaneously tracks facial expressions, hand shapes, and body posture
- Automatic frame padding for varied-length sign sequences
- Real-time keypoint detection with face, pose, and hand landmark tracking
- Intelligent Segmentation Innovation: Algorithm accurately detects boundaries between different signs in video streams

System Level Results

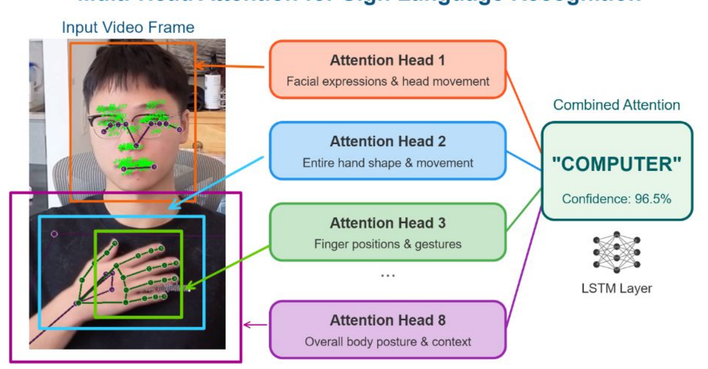
- **Browser ASL Translation**: Successfully developed a web-based pipeline that converts sign language videos into words using deep learning.
- **Efficient Cloud-Based Deployment:** Optimized critical scripts in the ML pipeline to accelerate processing on AWS, reducing inference time and improving scalability
- **Separation of Compute Tasks**: Separated tasks across AWS EC2 instances, using lightweight instances for frontend interactions and high-performance instances for compute-heavy ML related inference.
- **Seamless Frontend-Backend Communication**: Established seamless communication between a React-based frontend and Django backend, enabling quick processing and presentation of translations.
- **OpenAI Integration**: Integrated OpenAI services to refine raw model predictions into meaningful



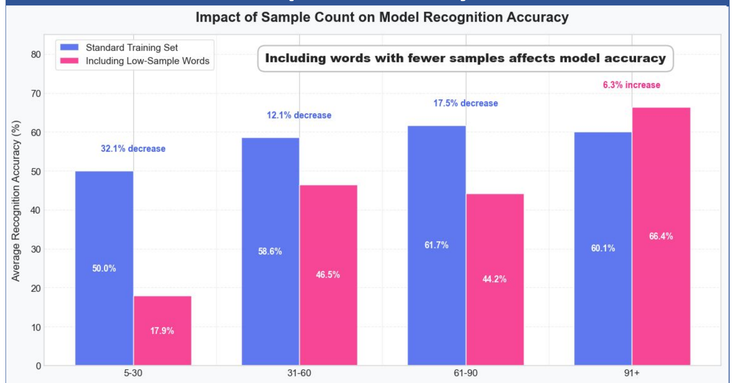
Backend system block diagram



Model attention architecture



Sample size analysis



Sample Group	Standard Words	Standard Samples	Low-Sample Words	Low-Sample Samples
5-30	3 words	avg 30.0 samples	7 words	avg 11.1 samples
31-60	76 words	avg 45.4 samples	77 words	avg 45.3 samples
61-90	7 words	avg 61.3 samples	7 words	avg 61.3 samples

Performance Highlights

Key Performance Metrics:

- **60.6%** accuracy across **100** ASL vocabulary words
- **95%+** accuracy on words with sufficient training data (91+ samples)
- Ultra-fast model inference: <**15ms** per word prediction
- High confidence predictions (up to **0.99**) for well-trained words

Challenges

1. Data Quality & Balance
   • Extreme class imbalance (some signs <8 samples) and inconsistent labeling
   • Solution: Statistical quality assessment and targeted data curation

2. Real-World Generalization
   • 90%+ training accuracy but lower real-world performance
   • Solution: Extensive data augmentation (temporal resampling, CutMix)

3. Continuous Recognition
   • Accurate boundary detection between flowing signs
   • Solution: Multi-stage intelligent segmentation algorithm

4. Computational Performance
   • MediaPipe keypoint extraction bottlenecks (~7-9s processing time)
   • Solution: Asynchronous inference reducing model time to <15ms

5. "Butterfly Effect"
   • Adding insufficient samples degraded performance across all words
   • Solution: Strategic filtering and incremental validation protocols

Conclusion and Future Work

We developed a sign language translation system to enhance communication for the Deaf and Hard-of-Hearing community. Using deep learning, our solution converts sign language gestures into text, refined through extensive design, implementation, and testing.

If we had more time, we would integrate database services to enhance user experience and efficiency.

© POSTER TEMPLATE BY GENGRAPHICS - 1.800.790.4001 - WWW.GENGRAPHICS.COM