Evaluating Semantic Context and Non-Manual Information for American Sign Language Recognition

Alicia Richard[1], Stefan Andrei[2]
[1]University of Texas at Austin
[2]Center for Computing, Education, and Instruction
[3]Center for Human-Machine Systems, Cleveland State University
[1]a.richard1@csuohio.edu, [2]s.andrei@csuohio.edu, [3]engineering@csuohio.edu

OBJECTIVE

Evaluate whether semantic context and non-manual information can improve a vision-based ASL recognition pipeline designed as a proof of concept for conversational applications.

METHODS

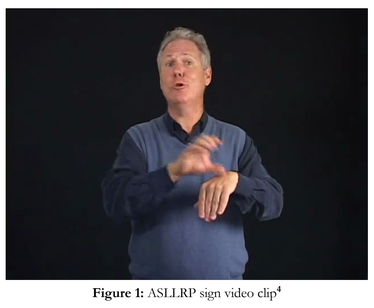


Figure 1: ASLLRP sign video clip[4]

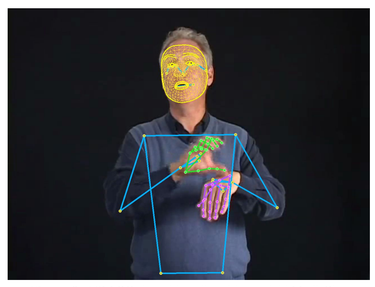


Figure 2: ASLLRP sign video clip with overlaid MediaPipe landmarks[4]

1 Dataset

ASL videos obtained from the American Sign Language Linguistic Research Project (ASLLRP)[4].



Figure 3: ASLLRP mp4 file

2 Landmark and Feature Extraction

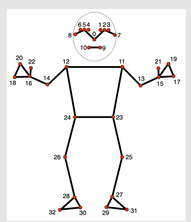


Figure 4: MediaPipe pose landmarker model[5]

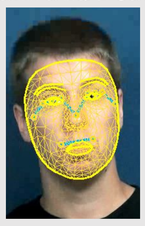


Figure 5: MediaPipe face mesh landmarker model

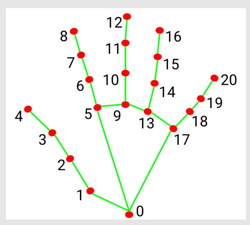


Figure 6: MediaPipe hand landmarker model[2]

3 Manual Sign Recognition

1D CNN classifies manual signs from extracted visual features. Outputs top-5 candidate signs.

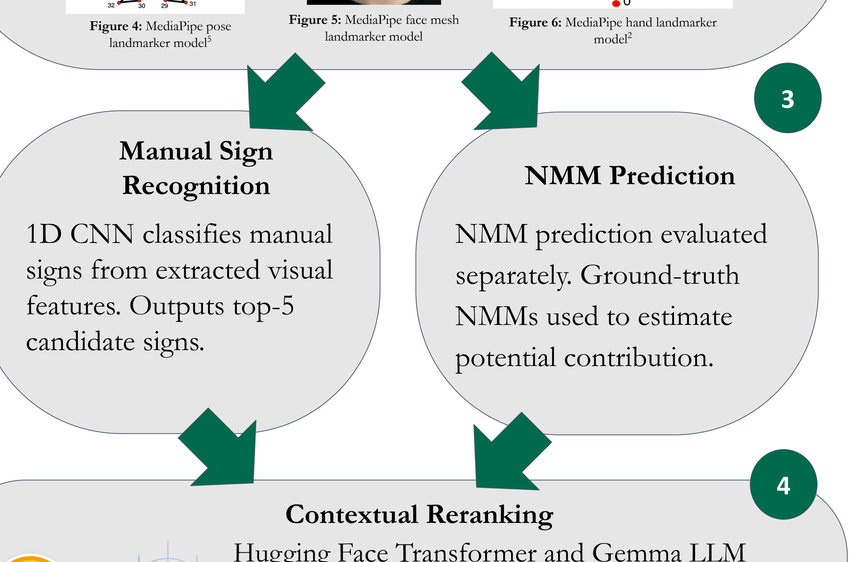
3 NMM Prediction

NMM prediction evaluated separately. Ground-truth NMMs used to estimate potential contribution.

4 Contextual Reranking

Hugging Face Transformer and Gemma LLM rerank uncertain tokens.

5 Final Gloss Sequence



Evaluation

- 73 classes with 230 examples grouped by source video
- Token accuracy, top-five oracle accuracy, and sentence exact match used as performance metrics

RESULTS

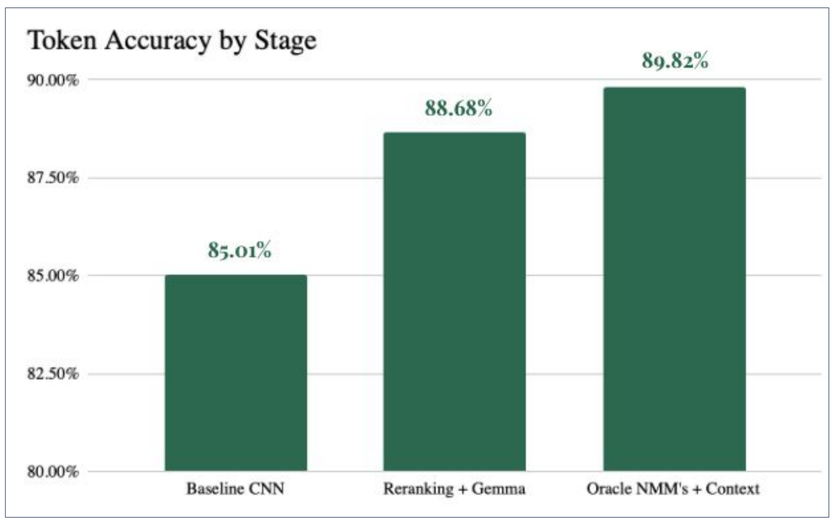


Figure 8: Exact sentence prediction accuracy, improved by context-awareness

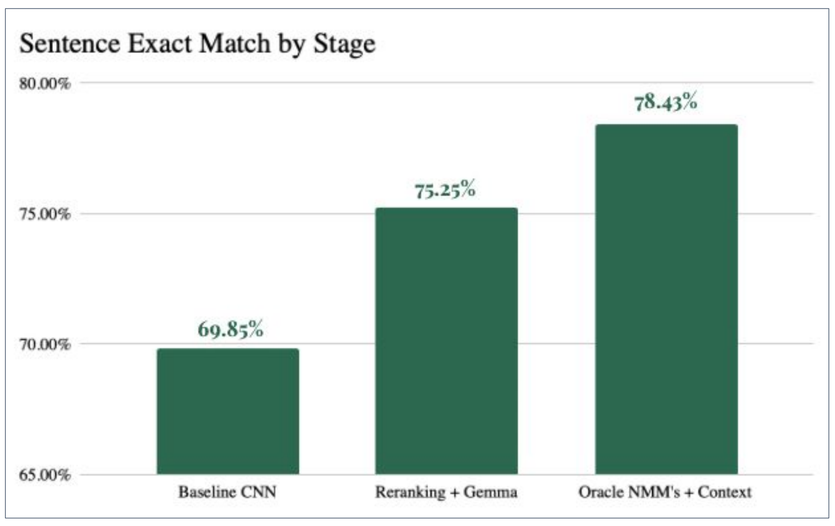


Figure 7: Token prediction accuracy, improved by context-awareness

Metric	Baseline CNN	Contextual Reranking	Oracle NMM's + Context*
Token Accuracy	85.01%	88.68%	89.82%
Sentence Exact Match	69.85%	75.25%	78.43%

Figure 9: Table summary of results

**Ground-truth NMM labels; represents an oracle condition.*

Reranking + Gemma

+3.67 percentage points token accuracy gain | +5.40 percentage points sentence exact match gain

vs Baseline CNN

Oracle NMM's + Context

+4.81 percentage points token accuracy gain | +8.58 percentage points sentence exact match gain

vs Baseline CNN

Conversational proof of concept

Input glossed utterance → English interpretation → English response + generated response gloss

Prompt-based prototype; not formally evaluated.

DISCUSSION AND FUTURE WORK

Semantic context improved both token-level recognition and complete-sentence accuracy, demonstrating the value of information beyond isolated visual signs. Providing ground-truth NMMs produced additional gains, particularly in sentence exact match, suggesting that reliable non-manual information could further improve recognition. However, NMM prediction was not sufficiently reliable for integration into the final pipeline, highlighting an important challenge for linguistically complete ASL recognition. Future work should improve NMM recognition and extend evaluation to continuous signing and low-resource signs. The pipeline should also be evaluated with Deaf and Hard-of-Hearing users before integration into signing avatars or other conversational technologies.

REFERENCES

1. Gemma 4 model overview | google AI for developers. (2026a, July 8). https://ai.google.dev/gemma/docs/core
2. Hand landmarks detection guide | google ai edge | google for developers. (2026b, May 28). https://developers.google.com/edge/mediapipe/solutions/vision/hand_landmarker
3. Mediapipe Solutions Guide | google ai edge | google for developers. (2026b, May 28). https://developers.google.com/edge/mediapipe/solutions/guide
4. Neidle, C., & Metaxas, D. (2025, June 28). *American Sign Language Linguistic Research Project (ASLLRP) Data Access Interface (DAI)*. American Sign Language Linguistic Research Project (ASLLRP) Sign Bank. http://dev.dai.cs.rutgers.edu/dai/s/index?redirect=dai
5. Pose Landmark Detection Guide | Google AI Edge | Google for developers. (2026d, May 28). https://developers.google.com/edge/mediapipe/solutions/vision/pose_landmarker
6. Sentence-transformers/all-mpnet-base-V2 · hugging face. (n.d.). https://huggingface.co/sentence-transformers/all-mpnet-base-v2