Supporting Sign Language Video Comprehension through In-Situ Automatic Segmentation and Sign Lookup Tool

Chaelin Kim
ckim13@tulane.edu
Tulane University
New Orleans, Louisiana, United States

Saad Hassan
saadhassan@tulane.edu
Tulane University
New Orleans, Louisiana, United States

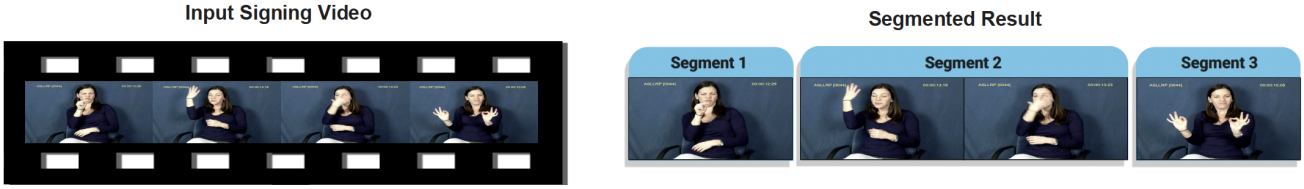


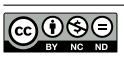
Figure 1: A conceptual overview of how automatic segmentation breaks continuous signing video into sign-level units, helping learners isolate unfamiliar signs, examine them individually, and access candidate matches without leaving the video context.

Abstract

Sign language learners often closely watch signing videos to develop receptive skills. However, they encounter unfamiliar signs, especially early on when they are still developing their vocabulary, and later due to challenges in understanding signs in continuous videos because of coarticulation. Existing tools, such as feature-based dictionaries, often require learners to identify the linguistic properties of a sign to search for it, which can be cumbersome. Prior work has proposed interfaces to support in-situ sign lookup, but end-to-end systems have not yet been disseminated or demonstrated. We demonstrate a fully automatic video comprehension prototype system that combines state-of-the-art sign language segmentation with a fine-tuned spatiotemporal recognition network. The underlying AI pipeline ingests a continuous signing video, applies a linguistically motivated BIO-tagging segmentation framework to locate boundaries, and routes individual segments into a fine-tuned Spatial-Temporal Graph Convolutional Network (ST-GCN) to retrieve the closest gloss matches. We present the system design and demonstration plan. We also discuss the system's current limitations and future research directions.

CCS Concepts

• **Human-centered computing** → *Accessibility technologies*; **User interface programming**.



ASSETS '26, Vila Nova de Gaia, Portugal

ACM ISBN 979-8-4007-2521-0/2026/10
https://doi.org/10.1145/3797867.3841227

Keywords

American Sign Language, Language Learning, Sign Language, Sign Language Learning, Artificial Intelligence, Sign Language Recognition, Sign Language Segmentation, AI Tools

ACM Reference Format:
Chaelin Kim and Saad Hassan. 2026. Supporting Sign Language Video Comprehension through In-Situ Automatic Segmentation and Sign Lookup Tool. In *The 28th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '26), October 25–28, 2026, Vila Nova de Gaia, Portugal.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3797867.3841227

1 INTRODUCTION

Sign languages, such as American Sign Language (ASL), are widely regarded as among the most challenging languages to learn [16], in part because they are expressed in a three-dimensional visual-spatial modality and employ grammatical mechanisms that differ fundamentally from those of spoken languages. Signs are produced through combinations of handshape, movement, palm orientation, spatial location, and nonmanual markers (e.g., facial expressions, head movements, and body posture), and their performance within continuous signing utterances can be shaped by the surrounding linguistic context [33].

Videos are commonly used in ASL learning, partly because learners may have limited access to people with whom they can practice, but also because videos are a central component of many ASL curricula [1]. However, a major bottleneck in video-based learning is the limited availability and poor usability of video comprehension tools. Most existing technological resources for ASL learning focus primarily on basic vocabulary and rarely extend beyond helping learners recognize isolated signs in their "citation" form[1]. Existing commercial and research-produced tools are typically designed either as vocabulary-learning games (e.g., [3, 19, 31]) or as feature-based dictionaries (e.g., [6, 20]) that allow learners to look

[1]In sign linguistics, a sign in "citation form" refers to its standard or dictionary form, produced in isolation rather than within a sentence or natural discourse [33].

up unfamiliar signs using their linguistic properties. Yet, identifying signs by linguistic properties can be especially challenging for beginner and intermediate sign language learners [13]. Even if learners attempt to infer these features from context, they must completely leave their video-viewing environment to perform a search, severely disrupting the comprehension process. Currently, no commercially available tools provide fluid, in-situ lookup or comprehension support for ASL video environments.

Fortunately, recent advances in sign segmentation [24, 27, 32] and recognition [8, 41] now make it possible to automate this supportive workflow. In particular, segmentation [24] has the potential to provide learners with video segments and the closest matching signs for each segment.

We will demonstrate an automatic, end-to-end ASL video comprehension prototype system. Building on prior studies that employed a Wizard-of-Oz method to investigate design parameters [12], this tool was originally developed and released for research purposes as part of a study on AI-based ASL video comprehension [17][2]. The system incorporates automatic sign language segmentation functionality, and its underlying AI pipeline combines state-of-the-art sign language segmentation with a fine-tuned recognition model that operates on the segmented outputs. The goal of this demonstration is to showcase how automatic segmentation and sign lookup can support learners and to gather informal feedback on the interaction design and performance of the underlying systems from ASSETS attendees, including d/Deaf, Hard of Hearing, and hearing signers and learners.

2 Related Work

2.1 Isolated Sign Language Recognition

Early sign recognition used handcrafted features and small datasets [23, 40] and relied on SVMs or HMMs [30]. The shift toward deep learning introduced 2D-CNNs and RNNs, which were gradually replaced by 3D-CNNs and Transformers capturing spatiotemporal dependencies [24, 38]. Concurrently, training corpora evolved into large-scale datasets like *MS-ASL* [15], *WLASL* [35], and *ASL Citizen* [8], enabling highly generalizable architectures like the Sign Pose Transformer [4]. However, these models mostly focus on signs in their isolated "citation form." To bridge this, Zhou et al. [41] recently coupled Graph Convolutional Networks (GCNs) with temporal boundary detection, improving accuracy for pre-segmented signs.

2.2 Sign Language Segmentation

Continuous Sign Language Recognition (CSLR) identifies successive signs without prior boundary knowledge [2], but remains challenging due to coarticulation [29]. Even with emerging foundation models (e.g., SignGemma [11], SignX [10]), learners heavily benefit from analyzing shorter segments. Thus, sign segmentation—breaking continuous videos into individual glosses—is critical. Early methods used pseudo-labelling and changepoint detection to refine boundaries [27]. Recently, Moryossef et al. [24] proposed a *linguistically motivated segmentation* framework using a BIO-tagging scheme, optical-flow prosodic features, and 3D hand normalization to detect precise boundaries, yielding the most consistent cross-linguistic results [32].

This demonstration integrates these advances by fine-tuning a recognition model adapted for both citation forms and automated segments [8, 41] using robust datasets [8, 25, 26], while leveraging the linguistically motivated pipeline [24] for our user interface.

[2]Prototype system: https://github.com/chaelin0722/asl-video-lookup-tool

Table 1: Recognition performance of the fine-tuned ST-GCN model on the ASLLRP dataset [25] (mean ± standard deviation).

Top-1	Top-5	Top-10	Top-20
0.33 ± 0.30	0.44 ± 0.38	0.48 ± 0.40	0.51 ± 0.42

3 AI Pipeline Architecture

We describe the processing pipeline in Figure 2 and present an overview in this section. For further details, please refer to prior work and software documentation on the system components [8, 17, 18, 24]. Notably, Kim et al. [18] provide a summary of both the AI pipeline and the interface design.

For *automatic segmentation*, the system uses the linguistically motivated segmentation model proposed in [24]. This model adopts a BIO-tagging scheme (Beginning–Inside–Outside) to identify the start and end points of each sign in continuous signing. The approach is grounded in linguistic cues such as prosody, motion, and handshape changes, and jointly models both sign- and phrase-level boundaries using pose-based features extracted from the signer's body and hands. Because it relies on motion and structural patterns rather than language-specific gloss annotations, the model is expected to generalize well across different sign languages. For the *isolated sign recognition* model, the system uses the Spatial-Temporal Graph Convolutional Network (ST-GCN) architecture [8]. ST-GCN is a pose-based model that provides competitive performance relative to other state-of-the-art models while enabling faster inference, making it well suited for integration into a real-time demonstration. The model, initialized with pretrained weights from the ASL Citizen dataset [8], is fine-tuned on a dataset of 991 glosses for 200 epochs using Stochastic Gradient Descent (SGD) [28] with a learning rate of 1×10^{-1}. The dataset exhibits substantial class imbalance, with gloss instances ranging from 1 to 289 per class. To address this imbalance, the training procedure employs a weighted random sampler that assigns higher sampling probabilities to underrepresented glosses [17].

Following the overall model pipeline described in [8], the system extracts 27 keypoints from the hands and face using MediaPipe Holistic [22], then center-scales and normalizes them. Data augmentation techniques, including random rotation and shear transformations on pose sequences, are applied as described in the original paper [8]. For optimization, the system uses cross-entropy loss with a cosine annealing learning rate scheduler [21] (T_{max} = 150). The encoder is initialized with pretrained weights, and the classifier is reinitialized using Xavier initialization [7]. We evaluate the fine-tuned recognition model on the ASLLRP dataset [25]. Table 1 summarizes recognition performance.

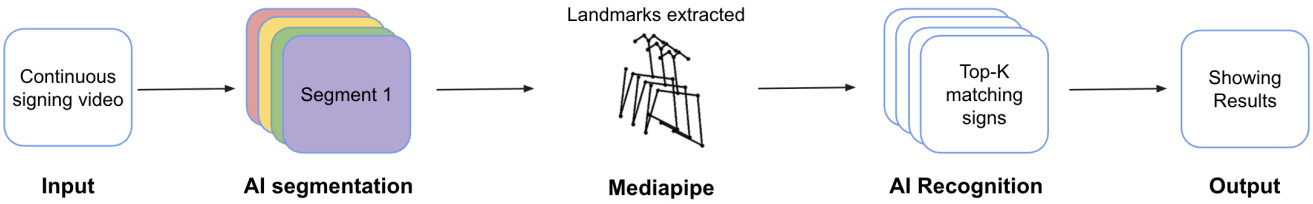


Figure 2: System architecture for automated ASL video comprehension. The system segments a continuous signing video into individual signs, extracts body and hand landmarks using MediaPipe, identifies the top-K matching signs, and presents the results to support learners' comprehension.

4 User Interface and Demo Plan

The demonstration will run on a 14-inch MacBook Pro with an M4 chip, which provides sufficient GPU compute power to support our models. We will provide attendees who may have a range of experience with ASL with a selection of sample videos at four different levels, pre-evaluated to ensure reasonable coverage of signs, which they can upload on a separate tablet. Next, we will familiarize them with the different components of the user interface with the help of the labeled diagrams in the accompanying poster.

As shown in Figure 3, the web-based system consists of five main components: a "`video player screen`", "`transcription panel`", "`search results panel`", "`video timeline`", and "`span selection functionality`". To support additional interactions implemented in the codebase [17], the system integrates four user-interface elements: a "`Segment and Search`" button, a "`Search`" button, and "`Upload Video`" and "`Recording`" buttons.

We will then encourage attendees to use the "`Segment and Search`" button, which allows them to trigger automatic segmentation and recognition on the segmentation output. After selecting a span on the video timeline using the adjustable span selector playhead by dragging the start and end markers, attendees can trigger the segmentation system to divide the selection into individual signs. This selection can include the entire duration of the video or a smaller segment that the user wants to analyze, so we will allow attendees to explore freely.

Once attendees click "`Segment and Search`" and processing is complete, a confirmation message will appear: "`Processing done, showing 20 results between selected start time–end time.`" Each automatically detected sub-segment is displayed in the navigation bar in chronological order, and selecting a sub-segment highlights the corresponding region on the timeline. This will allow attendees to quickly assess recognition quality and compare retrieved signs to their target signs without leaving the video context. The retrieved results are shown in the "`Search Results`" panel, where each sign is represented as a looping GIF along with a "Gloss" label[3] and a confidence score. Confidence levels are generated based on the system's confidence and are grouped into three ranges: Low (0–33%), Medium (34–66%), and High (67–100%). Finally, we will give attendees an opportunity to provide informal feedback through a survey and engage in discussion about future directions.

[3]An ASL gloss is a written label, typically an English word in capital letters, used to represent an ASL sign. Glossing serves as an interlinear notation system for representing ASL's visual-spatial grammar and structure [34].

5 Conference Discussion Opportunities and Future Directions

There are several opportunities to improve the system's AI models and interface, which we hope to discuss with ASSETS attendees. We outline some of these discussion points and future directions below.

As prior interdisciplinary workshops by Bragg et al. [5] and others have shown, the availability of well-annotated large datasets with diverse signers remains one of the biggest constraints in sign language understanding research. Even when such datasets are disseminated, they may not be publicly available or sufficiently representative of the diversity of signing practices. In the context of our system design, there is also a need for more annotated continuous-signing video data with sign-level annotations, which remains limited. Recent datasets, such as YouTube-ASL [37], are promising but will require significant effort from future researchers to annotate large corpora of videos using English captions.

As noted in the original work [24], the segmentation model was initially trained on German Sign Language (Deutsche Gebärdensprache) data rather than ASL data. Fortunately, recent methods, including the approach used in prior work [24], leverage centroid representations across multiple examples to stabilize performance [9] and can still achieve reasonable performance across different sign languages. However, future researchers could further improve segmentation by leveraging labeled continuous ASL datasets [26], or comparable datasets for other sign languages of interest, to fine-tune sign language-specific segmentation architectures. Future work can also investigate the use of large vision language models, e.g., Google Gemma [36] or Qwen2-VL [39], with prompts for automatic segmentation tasks.

Future researchers can also implement several UI enhancements to further lower the cognitive workload for independent learners. The current system performs well with a single signer. However, many videos online include multiple signers, such as conversations between two signers, or present sign language in only a small portion of the screen, such as through a sign language interpreter. This creates another opportunity to explore the design of *spatial* video segmentation, as opposed to only *temporal* segmentation. Future researchers can investigate this interaction.

Another limitation of the current interface is that users cannot visualize the detected segments within the timeline. Future work could provide this visualization to enable users to evaluate the detected segments based on their own understanding and adjust the segment boundaries as needed. Although the exact design should be

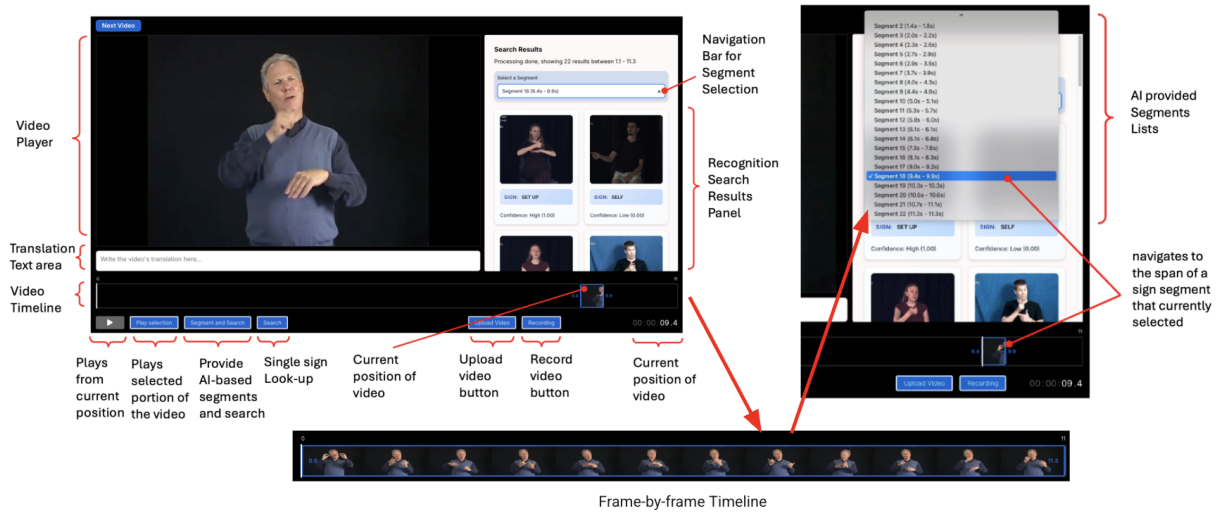


Figure 3: A scenario-based walkthrough of the learner interface. Users upload a signing video and activate "Segment & Search." The system automatically segments continuous signing into individual sign units and displays candidate recognition results for each segment in the results panel.

investigated, one idea is that the timeline bar could also be indexed with detected sign boundaries. By hovering over specific segments on the timeline, users could trigger immediate pop-ups displaying the top-matching glosses and confidence metrics, as proposed in prior research [14]. We hope the demo will surface similarly creative ideas for future investigation.

6 Conclusion

We plan to demonstrate a fully automated, end-to-end ASL video comprehension prototype that streamlines vocabulary lookup for sign language learners. The system utilizes an AI pipeline that automatically segments continuous signing videos and retrieves the closest matching sign glosses using an ST-GCN recognition model. The system provides top-k matching gloss-labeled GIFs directly within the learner's video-viewing environment. We hope to encourage researchers and practitioners to build such tools to support autonomous ASL learning and to discuss future directions.

Acknowledgments

The authors acknowledge the use of OpenAI's ChatGPT and Google's Gemini for text refinement. All text was reviewed to ensure it accurately reflected the authors' intended meaning. The authors acknowledge funding from the Louisiana Board of Regents Research Competitiveness Subprogram (RCS) award 093A-24.

References

[1] Jodie M Ackerman, Ju-Lee A Wolsey, M Diane Clark, et al. 2018. Locations of L2/Ln Sign Language Pedagogy. *Creative Education* 9, 13 (2018), 2037.

[2] Sarah Alyami, Hamzah Luqman, and Mohammad Hammoudeh. 2024. Reviewing 25 years of continuous sign language recognition research: Advances, challenges, and prospects. *Information Processing & Management* 61, 5 (2024), 103774.

[3] Dhruva Bansal, Prerna Ravi, Matthew So, Pranay Agrawal, Ishan Chadha, Ganesh Murugappan, and Colby Duke. 2021. CopyCat: Using Sign Language Recognition to Help Deaf Children Acquire Language Skills. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 481, 10 pages. doi:10.1145/3411763.3451523

[4] Matyáš Boháček and Marek Hrúz. 2022. Sign Pose-Based Transformer for Word-Level Sign Language Recognition. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) Workshops*. 182–191.

[5] Danielle Bragg, Oscar Koller, Mary Bellard, Larwan Berke, Patrick Boudreault, Annelies Braffort, Naomi Caselli, Matt Huenerfauth, Hernisa Kacorri, Tessa Verhoef, Christian Vogler, and Meredith Ringel Morris. 2019. Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility* (Pittsburgh, PA, USA) *(ASSETS '19)*. Association for Computing Machinery, New York, NY, USA, 16–31. doi:10.1145/3308561.3353774

[6] Danielle Bragg, Kyle Rector, and Richard E Ladner. 2015. A user-powered American Sign Language dictionary. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 1837–1848.

[7] Leonid Datta. 2020. A survey on activation functions and their relation with xavier and he normal initialization. *arXiv preprint arXiv:2004.06632* (2020).

[8] Aashaka Desai, Lauren Berger, Fyodor Minakov, Nessa Milano, Chinmay Singh, Kriston Pumphrey, Richard Ladner, Hal Daumé III, Alex X Lu, Naomi Caselli, et al. 2023. ASL citizen: a community-sourced dataset for advancing isolated sign language recognition. *Advances in Neural Information Processing Systems* 36 (2023), 76893–76907.

[9] Aashaka Desai, Daniela Massiceti, Richard Ladner, Hal Daumé III, Danielle Bragg, and Alex X Lu. 2025. Investigating Dictionary Expansion for Video-based Sign Language Dictionaries. In *Findings of the Association for Computational Linguistics: EMNLP 2025*. 22826–22841.

[10] Sen Fang, Chunyu Sui, Hongwei Yi, Carol Neidle, and Dimitris N Metaxas. 2025. SignX: The Foundation Model for Sign Recognition. *arXiv preprint arXiv:2504.16315* (2025).

[11] Google. 2025. SignGemma: an upcoming open model for sign-to-speech translation announced at Google I/O 2025. https://blog.google/technology/developers/google-ai-developer-updates-io-2025/. Official Google I/O 2025 announcement of SignGemma (Gemma family).

[12] Saad Hassan, Akhter Al Amin, Caluã de Lacerda Pataca, Diego Navarro, Alexis Gordon, Sooyeon Lee, and Matt Huenerfauth. 2022. Support in the Moment: Benefits and use of video-span selection and search for sign-language video comprehension among ASL learners. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–14.

[13] Saad Hassan, Akhter Al Amin, Alexis Gordon, Sooyeon Lee, and Matt Huenerfauth. 2022. Design and Evaluation of Hybrid Search for American Sign Language to English Dictionaries: Making the Most of Imperfect Sign Recognition. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 195, 13 pages. doi:10.1145/3491102.3501986

[14] Saad Hassan, Caluã de Lacerda Pataca, Akhter Al Amin, Laleh Nourian, Diego Navarro, Sooyeon Lee, Alexis Gordon, Matthew Watkins, Garreth W. Tigwell, and Matt Huenerfauth. 2024. Exploring the Benefits and Applications of Video-Span Selection and Search for Real-Time Support in Sign Language Video Comprehension among ASL Learners. *ACM Trans. Access. Comput.* 17, 3, Article 14 (Oct. 2024), 35 pages. doi:10.1145/3690647

[15] Hamid Reza Vaezi Joze and Oscar Koller. 2018. Ms-asl: A large-scale data set and benchmark for understanding american sign language. *arXiv preprint arXiv:1812.01053* (2018).

[16] Mike Kemp. 1998. Why is learning American Sign Language a challenge? *American Annals of the Deaf* (1998), 255–259.

[17] Chaelin Kim, Denise Crochet, Maty Bohacek, and Saad Hassan. 2026. ASL Video Lookup Tool. https://github.com/chaelin0722/asl-video-lookup-tool. GitHub repository. Accessed: 2026-07-01.

[18] Chaelin Kim, Denise Crochet, Maty Bohacek, and Saad Hassan. 2026. Design and Evaluation of an AI-Based American Sign Language Video Comprehension Tool: Exploring the Benefits and Use of Automatic Segmentation and Sign Lookup. In *The 28th International ACM SIGACCESS Conference on Computers and Accessibility* (Vila Nova de Gaia, Portugal) *(ASSETS '26)*. Association for Computing Machinery, New York, NY, USA, 19 pages. doi:10.1145/3797867.3828986

[19] Chaelin Kim, Cameron McLaren, Madhangi Krishnan, Nikhil Modayur, and Saad Hassan. 2025. Signing for Care: A Demo and Initial Evaluation of an American Sign Language Learning Tool for Emergency Medical Service Providers. In *Proceedings of the 27th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–6.

[20] J. Lapiak. 2021. Handspeak. https://www.handspeak.com/

[21] Ilya Loshchilov and Frank Hutter. 2017. SGDR: Stochastic Gradient Descent with Warm Restarts. In *International Conference on Learning Representations*. https://openreview.net/forum?id=Skq89Scxx

[22] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, et al. 2019. MediaPipe: A framework for building perception pipelines. *arXiv preprint arXiv:1906.08172* (2019).

[23] Aleix M Martínez, Ronnie B Wilbur, Robin Shay, and Avinash C Kak. 2002. Purdue RVL-SLLL ASL database for automatic recognition of American Sign Language. In *Proceedings. Fourth IEEE International Conference on Multimodal Interfaces*. IEEE, 167–172.

[24] Amit Moryossef, Zifan Jiang, Mathias Müller, Sarah Ebling, Yoav Goldberg, Juan Pino, and Kalika Bali. 2023. Linguistically Motivated Sign Language Segmentation. In *Findings of the Association for Computational Linguistics: EMNLP 2023*. Association for Computational Linguistics, Singapore, 12703–12724. doi:10.18653/v1/2023.findings-emnlp.846

[25] Carol Neidle, Augustine Opoku, and Dimitris Metaxas. 2022. Asl video corpora & sign bank: Resources available through the american sign language linguistic research project (asllrp). *arXiv preprint arXiv:2201.07899* (2022).

[26] Carol Neidle, Ashwin Thangali, and Stan Sclaroff. 2012. Challenges in development of the american sign language lexicon video dataset (asllvd) corpus. In *5th workshop on the representation and processing of sign languages: interactions between corpus and Lexicon, LREC*.

[27] Katrin Renz, Nicolaj C. Stache, Neil Fox, Gul Varol, and Samuel Albanie. 2021. Sign Segmentation With Changepoint-Modulated Pseudo-Labelling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. 3403–3412.

[28] Herbert Robbins and Sutton Monro. 1951. A stochastic approximation method. *The annals of mathematical statistics* (1951), 400–407.

[29] Jérémie Segouat. 2009. A study of sign language coarticulation. *SIGACCESS Access. Comput.* 93 (Jan. 2009), 31–38. doi:10.1145/1531930.1531935

[30] Ozge Mercanoglu Sincan, Julio C. S. Jacques Junior, Sergio Escalera, and Hacer Yalim Keles. 2021. ChaLearn LAP Large Scale Signer Independent Isolated Sign Language Recognition Challenge: Design, Results and Future Research. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. 3472–3481.

[31] Thad Starner, Sean Forbes, Matthew So, David Martin, Rohit Sridhar, Gururaj Deshpande, Sam Sepah, Sahir Shahryar, Khushi Bhardwaj, Tyler Kwok, et al. 2023. Popsign ASL v1. 0: An isolated american sign language dataset collected via smartphones. *Advances in Neural Information Processing Systems* 36 (2023), 184–196.

[32] Ariel E Stassi, J Matías Di Martino, and Gregory Randall. 2025. A Brief Review and Analysis of Two Methods for Automatic Sign Language Segmentation. *Image Processing On Line* 15 (2025), 59–77.

[33] William C Stokoe. 1980. Sign language structure. *Annual review of anthropology* (1980), 365–390.

[34] Samuel J Supalla, Jody H Cripps, and Andrew PJ Byrne. 2017. Why American sign language gloss must matter. *American annals of the deaf* 161, 5 (2017), 540–551.

[35] Federico Tavella, Viktor Schlegel, Marta Romeo, Aphrodite Galata, and Angelo Cangelosi. 2022. WLASL-LEX: a dataset for recognising phonological properties in American Sign Language. *arXiv preprint arXiv:2203.06096* (2022).

[36] Gemma Team, Morgane Riviere, Shreya Pathak, Pier Giuseppe Sessa, Cassidy Hardin, Surya Bhupatiraju, Léonard Hussenot, Thomas Mesnard, Bobak Shahriari, Alexandre Ramé, et al. 2024. Gemma 2: Improving open language models at a practical size. *arXiv preprint arXiv:2408.00118* (2024).

[37] Dave Uthus, Garrett Tanzer, and Manfred Georg. 2023. Youtube-asl: A large-scale, open-domain american sign language-english parallel corpus. *Advances in Neural Information Processing Systems* 36 (2023), 29029–29047.

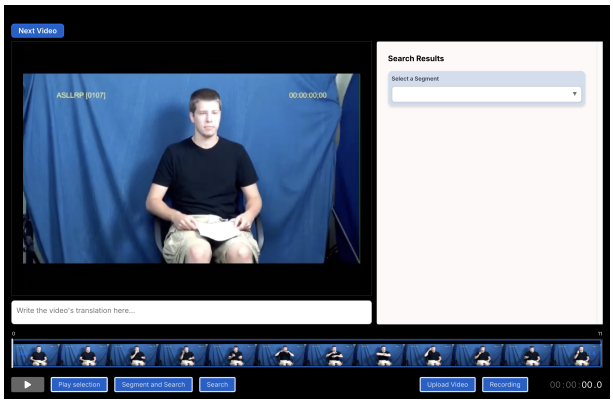
[38] Manuel Vazquez-Enriquez, Jose L. Alba-Castro, Laura Docio-Fernandez, and Eduardo Rodriguez-Banga. 2021. Isolated Sign Language Recognition With Multi-Scale Spatial-Temporal Graph Convolutional Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. 3462–3471.

[39] Peng Wang, Shuai Bai, Sinan Tan, Shijie Wang, Zhihao Fan, Jinze Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, et al. 2024. Qwen2-vl: Enhancing vision-language model's perception of the world at any resolution. *arXiv preprint arXiv:2409.12191* (2024).

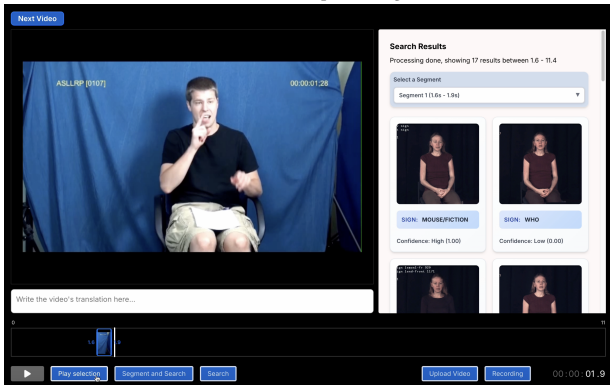
[40] Mahmoud M Zaki and Samir I Shaheen. 2011. Sign language recognition using a combination of new vision based features. *Pattern recognition letters* 32, 4 (2011), 572–577.

[41] Yang Zhou, Zhaoyang Xia, Yuxiao Chen, Carol Neidle, and Dimitris Metaxas. 2024. A Multimodal Spatio-Temporal GCN Model with Enhancements for Isolated Sign Recognition. In *Proceedings of the LREC-COLING 2024 11th Workshop on the Representation and Processing of Sign Languages: Evaluation of Sign Language Resources*, Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Julie A. Hochgesang, Johanna Mesch, and Marc Schulder (Eds.). ELRA Language Resources Association (ELRA) and the International Committee on Computational Linguistics (ICCL), Torino, Italy, 408–419. doi:10.63317/5jojxksciv3h

A The interface before and after executing a search



(a) Prior to AI processing



(b) After AI processing

Figure 4: System interface states (a) before and (b) after executing the AI pipeline.