As per UGC guidelines an electronic bar code is provided to secure your paper

International Journal for Modern Trends in Science and Technology
Volume 12, Issue 03, pages 74-80.
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol12issue03.html
DOI: https://doi.org/10.5281/zenodo.18880611

Sign Language Recognition for Inclusive Communication

Suneetha Sikile[1], S Naga Pavan Chandana Thayaru Manepalli[2], Gayatri Sowmya Kondapalli[2], Venkata Siva Prasad Deyyala[2], Taraka Venkat Dangeti[2]

Department of ECE, Godavari Global University, Rajamahendravaram, INDIA.
Department of ECE, Godavari Institute of Engineering & Technology (A), Rajamahendravaram, INDIA.

To Cite this Article
Suneetha Sikile, S Naga Pavan Chandana Thayaru Manepalli, Gayatri Sowmya Kondapalli, Venkata Siva Prasad Deyyala & Taraka Venkat Dangeti (2026). Sign Language Recognition for Inclusive Communication, 12(03), 74-80. https://doi.org/10.5281/zenodo.18880611

Article Info
Received: 28 January 2026; Revised: 26 February 2026; Accepted: 02 March 2026.

KEYWORDS

Sign Language Detection, Real-Time Gesture Recognition, Bidirectional Translation, YOLO Object Detection, Flask Web Application, Text-to-Sign Conversion, Learning Module, Search-Enabled Sign Phrases.

ABSTRACT

The proposed Sign Language Detection and Translation System is a bidirectional and real-time communication platform that aims to help people who do not know sign language and those who use it in the opposite direction, it aids communication by translating spoken or typed English words into corresponding sign representations. Moreover, the system provides a well-organized three-level learning module with video tutorials and multiple-choice questions to facilitate learning and accessibility. This module allows users to move from learning basic signs to having a conversation. In addition to this, there is a daily words module with an integrated search engine that keeps track of new signs when they are not available locally and that provides sign examples for the most common terms. The whole system is aimed at supporting and including the ones who need practical communication the most.

INTRODUCTION

For a considerable number of persons that are hard of hearing or deaf, sign language is the principal way of communication. Nonetheless, sign language is still very important but there are many people who do not have any sign language skills, which results in misunderstandings and social isolation. The technological link between signers and non-signers has been created by the recent inventions in computer vision and deep learning which enabled the development of intelligent systems that can recognize visual gestures. Research on real-time gesture recognition has shown that machine learning models, mainly convolutional and transformer-based architectures, can accurately recognize hand shapes, orientations, and movement patterns, thus shrinking linguistic gaps in diverse social and professional contexts [1].

Sign language recognition has made remarkable progress since the invention of fast and accurate object detection models like YOLO (You Only Look Once). YOLO's single-stage detection pipeline enables effective frame-by-frame gesture identification, which makes it suitable for real-time applications where low latency is required. Studies have shown that YOLO-based systems outperform many traditional approaches in recognizing both static and dynamic gestures in challenging environments with varying light or occlusions [4]. Besides, the advancement of sign language recognition technology has been supported by recent works using YOLOv5, YOLOv8, and YOLOv9 variants, which have exhibited not only higher accuracy but also robustness in discerning alphabet signs, frequent words, and fluid motion [6].

The new technologies strive to go far beyond gesture-to-text recognition and aim to help both the parties in communication. This method, on the one hand, gives non-signers the chance to communicate with sign language users, and on the other, it goes hand in hand with large-scale initiatives that advocate for inclusivity. It has been found in research that the use of gesture communication, speech processing, and natural language understanding together in the so-called multi-modal translation systems can result in greater accessibility in public places and educational institutions [3]. Not only that, but also bringing in the gesture-learning modules—like practice videos, quizzes, and guidance—non-signers can gradually learn the art of signing in an organized manner which, in a way, will support the automated translation technology [12].

The recognition of gestures and non-manual expressions in sign languages has received a considerable boost through the availability of extensive annotated sign language datasets and improved annotation tools. While deep learning models are still facing challenges on sign languages like hand-size differences, motion blur, noise, and personal quirks, they are still able to a greater extent build their performance on the generalization aspects of the models [21]. As the research goes towards the direction of sign language recognition being completely natural and translations being at the level of human comprehension, bidirectional systems of real-time communications have become very powerful tools for inclusion, communication, and even the improvement of life quality for those depending on the use of sign language [23].

RELATED WORKS

The use of deep learning and real-time object detection models have significantly increased the research areas of sign language recognition. Initially, the research focused on CNNs for the classification of static hand gestures which showed good results but were not suitable for real-time applications. Later on, as hardware and model optimization evolved, the researchers started using motion-based features to obtain better dynamic gesture recognition. The systems built on improved CNN architectures were good for isolated gesture recognition but did not perform well with continuous signing and variations in the environment like lighting or background noise [10]. The early studies pointed out the requirement of feature extraction that is robust for precise sign interpretation [5].

The arrival of YOLO-based architectures not only made the gesture detection but also detection of the hand regions in real-time through a single-stage processing revolutionized gesture detection. Studies, where YOLOv3, YOLOv4, and YOLOv5 were used, revealed a huge increase in speed and accuracy of detection, especially in noise or mess prone environments [9]. Further, studies employing YOLOv8 and YOLOv9 have enhanced gesture classification via better attention mechanisms and anchor-free detection strategies, thus making them compatible with real-time translation systems [4]. Support of these advancements provided the capturing of hand signs off video streams with no delay thus they were the pioneers of fast sign language interpretation platforms [6].

In recent times, there has been a development of systems that allow use of gestures beyond the recognition of only text. Various models are translating spoken languages or typed texts into sign equivalents basing on sign databases composed of sign images, pose keypoints, or animated signs. The combination of natural language processing (NLP) and YOLO-based gesture recognition made the translation between modalities less intrusive and more context-aware thus making the whole process more accessible to both users [3]. This also led to the integration of learning modules such as tutorials, quizzes, and structured lessons in the systems, which have particularly gained popularity among the individuals who want to learn basic or intermediate signing skills without formal education [12].

Building comprehensive, scalable, and adaptable sign language frameworks that can handle large vocabularies and continuous gestures is a key focus of modern studies. Usage of transfer learning and large annotated datasets, including Indian Sign Language (ISL) and American Sign Language (ASL), bring about the generalization across users and conditions significantly better [21]. Continuous sign language recognition—they say it is the hardest task—has been advancing by using hybrid architectures that fuse object detection with sequence modelling for sentence-level understanding [23]. All these developments symbolize the increasing demand of inclusive, real-time communication systems like that of the ordinary people's interaction.

EXISTING SYSTEM

Most of the current sign language recognition technologies are using conventional computer vision processes as their basic techniques for hand region identification. These processes include background subtraction, skin-color segmentation, and contour detection. The simplicity of these methods is one of their main advantages, yet they are still very sensitive to environmental conditions like lighting changes, the presence of other objects in the background, and different skin colors. This makes them unsuitable for real-world situations [10]. For a long time, the development of such systems was mainly focused on the classification of static gestures. Thus, they could not process or recognize continuous or dynamic signs, which ultimately limited their usage to full sign language communication. Furthermore, these models were often restricted to controlled environments with consistent backgrounds, resulting in reduced flexibility and scalability when trying to use them in natural settings [5].

Machine learning advancements made it possible for deep learning-based systems to bring about a radical change by incorporating CNNs into gesture classification which resulted in increased accuracy. While CNNs allowed better and more hardship tolerant feature extraction compared to the classical methods, they still could not manage real-time processing due to the high computational requirements. Therefore, many current models were limited to classifying only pre-captured images and thus not able to handle live video streams which considerably restricted their use in interactive communication [1]. In addition to that, most CNN-based systems were excessively dependent on large annotated datasets which were not available for many sign languages, including Indian Sign Language (ISL) and its regional variants, which made the efforts less fruitful [12].

The entry of object detection frameworks like YOLO was a huge step forward as it allowed for real-time gesture recognition, thus up to now, detection capabilities were made possible in a whole new way. Unfortunately, the majority of the YOLO-based sign language systems that exist today are still limited to only the signs of alphabets or small vocabularies, and that makes it impossible to do a complete language translation [6]. Some of the models cannot understand continuous gestures or do not support sign language at the sentence level due to the nonexistence of components that handle temporal modeling. Plus, current methods are usually one-way only, converting gestures into text but not vice versa, thus non-signers have a harder time getting involved in the process [4].

The systems that are already in place are not really targeting the novice learners and hence, their educational value is very low. The majority of the platforms offer only the gesture detection service without providing the accompanying structured learning modules, practice exercises, or guided tutorials to enable users to gradually learn sign language skills. Furthermore, the popular applications do not include features such as phrase libraries, integrated search functionality, or adaptive gesture suggestions, which consequently restrict the actors support in daily communication scenarios [21]. Thus, the current systems are not able to fully cater to the communication needs of both signers and non-signers in the practical context [23].

PROPOSED METHODOLOGY

The system that is being proposed comes with a cutting-edge, real-time, two-way sign language communication platform specifically made to mitigate the downfalls of current gesture recognition methods. The system is based on YOLO object detection machinery, and it properly identifies hand gestures coming from the live video input and renders them in English text. The web-based interface which uses Flask provides uninterrupted interaction by merging gesture recognition, text display, and speech synthesis into a single platform. This real-time conversion makes it possible for sign language users to converse with non-signers in an effortless manner, without the need of any external interpreters or controlled environments.

Along with this, the proposed system also has a reverse translation mechanism that converts spoken or typed English sentences into corresponding sign language gestures; this is because of the systematic embedding of speech-to-text processing, natural language understanding, and a limited sign gesture library. With this, the system enables non-signers to communicate in a more natural way with those who are dependent on sign language. It is the bidirectional feature that signifies a great enhancement compared to the conventional systems that are limited to gesture-to-text conversion, thus, opening the door to more communication scenarios where accessibility and inclusivity are the main virtues.

The main breakthrough in the suggested scheme is a structured learning module tailored for people who want to learn sign language. The module consists of three progressive levels that are arranged by difficulty, with each level providing videos, tasks for practice, and assessments in the form of multiple-choice questions. The interactive method makes the users move slowly but surely from the stage of learning the basic gestures of the alphabet to the stage of understanding words and having a signing conversation. The module constitutes an educational asset for the whole system and allows beginners, teachers, and the families of the deaf or hard-of-hearing individuals to learn at their own pace. Existing System.

Moreover, a Daily Phrases module that shows the most common phrases represented through sign language is added to the system. Users can access signs for words not found in the local dataset because of the integrated search engine API, which guarantees a complete phrase support. The proposed system integrates gesture recognition, translation, learning, and searchable sign resources to provide a total solution that not only empowers the signers and non-signers but also closes the communication gaps and enhances the real-world usability across a variety of settings.

Architecture Diagram

The system's design lays out a patchwork of a real-time communication framework that marries computer vision, deep learning, natural language processing, and a Flask-based web user interface. YOLO at the center detector does the job of filtering recognizing and classifying the signs made with hands consonant to the sign language through live input from the webcam. OpenCV is used to take each video frame, do some preprocessing, and then give it to the YOLO inference module wherein the box for gesture recognition and class labels are done instantly. The detected gesture is then matched with its equivalent English letter or word and shown on the web interface. On the other hand, a text-to-speech (TTS) engine is producing audible output of the interpreted text, thus allowing interaction between the signers and the non-signers to be smooth.

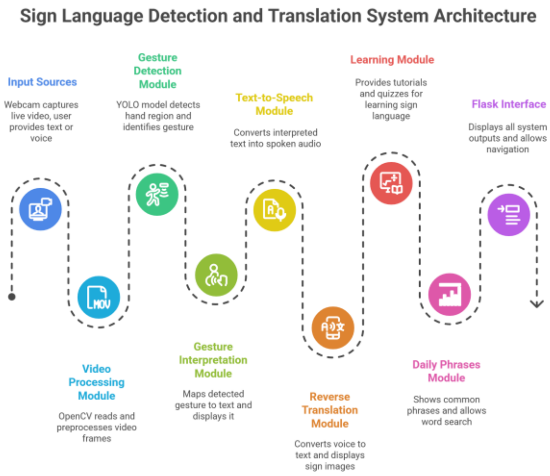

Figure 1: System architecture

The architecture not only recognizes gestures but also provides a means to turn translations back using speech-to-text APIs incorporated into a structured gesture library. The user can speak or write text and the system will show the corresponding sign language gesture and will do so either using images stored or through animations. The system's usability is augmented by the addition of other modules such as a Learning Module that provides videos and assessments according to the learner's level, and the Daily Phrases Module that has an embedded search API to dynamically retrieve even from external sources signs that are not familiar. Flask routes connect all the components thereby making the data flow between the frontend interface and backend processing units more efficient. This modular architecture enables the system to be easily adapted, expanded, and employed in real-world, multi-user situations.

ALGORITHM

Real-time sign language recognition is mainly based on the YOLO (You Only Look Once) object detection algorithm, which is the proposed system to the foremost. YOLO adopts a single-stage detection technique, wherein the input image is partitioned into grid cells, and each cell predicts the bounding boxes and class probabilities concurrently. This is in contrast to multi-stage methods, which divide the proposed region and classification tasks and consequently slow down the processing. A single pass through the neural network suffices for YOLO, thus achieving phenomenal speed. As a result, the system can handle uninterrupted video frames with practically no time lag, making it fit for real-time interaction. YOLO's anchor-free design in its latest iterations such as YOLOv8 and YOLOv9 not only improves the detection quickness but also the detection accuracy in difficult scenarios, particularly for small and complex hand regions that differ in shape, attitude and background conditions.

The system, beyond gesture detection, further is augmented by the integration of text-to-speech, speech-to-text, and sign gesture mapping algorithms which are supportive of the bidirectional translation. Audio input is handled through a speech recognition algorithm which transforms the audio input into text, on the other hand, the text-to-sign module is using predefined gesture datasets or dynamic search results to showcase the relevant sign visuals. The learning module employs simple adaptive algorithms to assess user responses via quizzes and monitor progress through levels. These algorithms together create a seamless and efficient pipeline that can translate gestures, produce audio outputs, and facilitate structured learning. The dual application of YOLO-based vision algorithms and NLP-driven processing guarantees great accuracy, slight delay, and resilient performance in everyday communication settings.

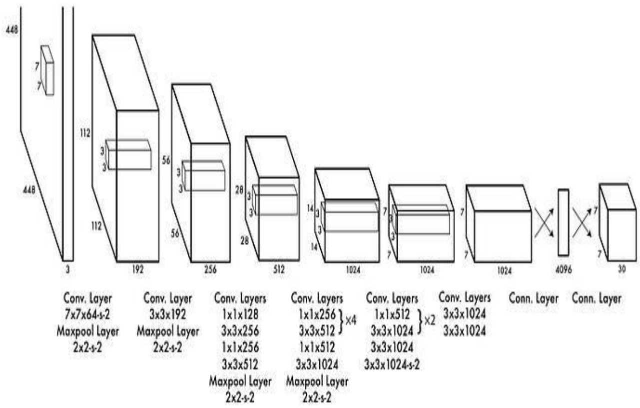

Figure 2: Architecture of yolo

RESULTS AND DISCUSSION

The YOLO-based sign language detection model was powerful enough to recognize gestures with almost perfection. The precision was 0.91 which meant that the prediction was almost always right and the recall was 0.89 which means that the system did not miss the majority of the signs that were actually there. The resulting F1-score of 0.90 is an indication of a well-balanced and reliable performance that is suitable for real-time applications where accuracy consistency is a critical factor.

In terms of detection accuracy, the model produced an mAP@0.5 of 0.92 which attests to the model's effectiveness in not only recognizing but also locating sign gestures at the standard evaluation threshold. The strong detection performance is a demonstration of the model's robustness and it opens up the possibility of using the model in various applications like sign language translation and learning. The results thus far indicate that the proposed system is not just accurate but also very fast in recognizing gestures for real-time communication.

Table 1: Performance metrics

Metric	Value
Precision	0.91
Recall	0.89
F1-Score	0.90
mAP@0.5	0.92

7.1 Translator Module Output

Figure 3: Speech-to-sign translation output showing the recognized letters for the word "hello".

7.2 Learning Module – Basic Level Output

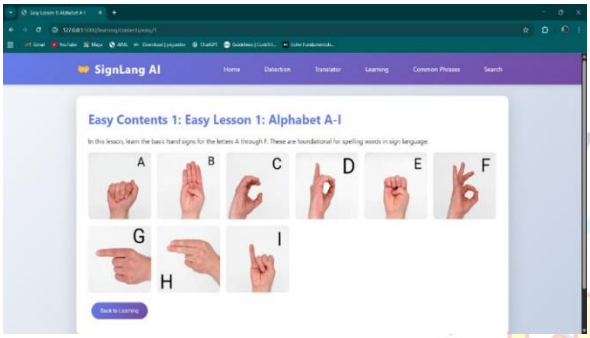

Figure 3: Learning module interface displaying alphabet signs from A to I.

7.3 Common Phrases Module Output

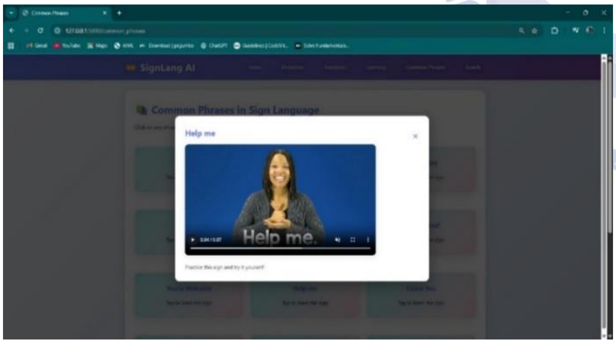

Figure 4: Demonstration video for the Common Phrases module.

7.4 Learning Module – Quiz Output

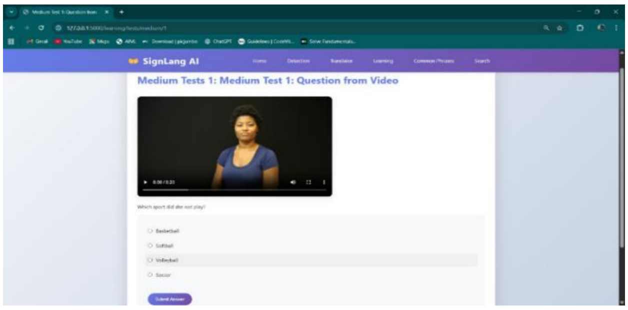

Figure 5: Assessment with multiple-choice answers

7.5 Real-Time Gesture Detection Output

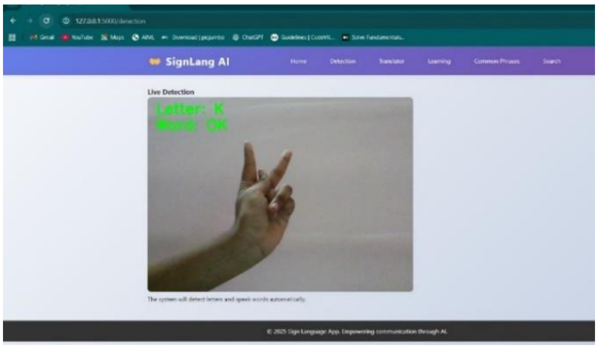

Figure 6: Real-time YOLO-based detection

Conclusion

The real-time sign language detection and translation system proposed above is a very effective way of bridging the gap between sign language users and non-signers since it brings together the very latest in computer vision, deep learning, and interactive learning techniques. The YOLO model specifically allows the system to perform very accurate hand gesture recognition and the model to produce plain text as well as on-the-fly translation of the spoken or typed words back into the corresponding sign representations. The learning modules, daily phrase demonstrations, and search-assisted phrase retrieval are the non-stop automated features of the system that makes it not only a communication tool but also an educational and skill development one.

The experiment shows that the model performed excellently as it achieved remarkable figures in terms of precision, recall, F1-score, and mAP, which confirmed that it is indeed a reliable tool for real-world applications. Furthermore, the easy-to-use interface and the very responsive modules are two features of the system that prove its capability for practical, continuous interaction. To sum up, the project has been able to create a platform that is inclusive, accessible, and scalable as it supports the seamless communication of the deaf and hard-of-hearing community. The dataset can be enlarged, and continuous sign recognition and multilingual support can be added in the future to improve the application further in terms of its versatility and impact.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] M. Alaftekin, I. Pacal, and K. Cicek, "Real-time sign language recognition based on YOLO algorithm," Neural Computing and Applications, 2024. https://link.springer.com/article/10.1007/s00521-024-09503-6

[2] R. Sreemathy, "Continuous word level sign language recognition using an expert system," Scientific Reports, 2023. . https://www.sciencedirect.com/science/article/pii/S2666307423000177

[3] B. Alsharif et al., "Real-Time American Sign Language Interpretation Using Deep Learning and Keypoint Tracking," Sensors, 2025. https://pmc.ncbi.nlm.nih.gov/articles/PMC11991612

[4] A. Imran, M. S. Hulikal, H. A. A. Gardi, "Real Time American Sign Language Detection Using YOLO-v9," arXiv preprint, 2024. https://arxiv.org/abs/2407.17950

[5] W. Jia, "SLR-YOLO: An improved YOLOv8 network for real-time sign language recognition," Journal of Information & Forensics Security, 2024. https://journals.sagepub.com/doi/full/10.3233/JIFS-235132

[6] M. Biyani and P. Doohan, "Real-Time Sign Language Recognition Using YOLOv5," 2024.https://www.semanticscholar.org/paper/Real-Time-Sign-Language-Recognition-Using-Yolov5-Biyani-Doohan

[7] U. Rastogi et al., "Advanced gesture recognition in Indian sign language using modified deep learning network and hybrid optimization," Scientific Reports, 2025. https://www.nature.com/articles/s41598-025-18496-8

[8] TechScience, "Recent Advances on Deep Learning for Sign Language Recognition: A Survey," CMES, 2024. https://www.techscience.com/CMES/v139n3/55626

[9] A. Mujahid et al., "Real-time hand gesture recognition based on deep-learning YOLOv3 model," Applied Sciences, 2021.https://www.mdpi.com/2076-3417/11/13/5900

[10] H. Fang et al., "Sign language recognition: A deep learning overview and research frontier," Journal of Ambient Intelligence and Humanized Computing, 2021. https://link.springer.com/article/10.1007/s12652-020-02897-4

[11] H. Y. A. Swasono et al., "YOLO11 and OpenCV Implementation for Phrase Recognition in Real-Time Hand Sign Language Videos," MALCOM Journal, 2025. https://www.journal.irpi.or.id/index.php/malcom/article/view/2130

[12] H. Wang, "Research of a Sign Language Translation System Based on Deep Learning," ResearchGate, 2019. https://www.researchgate.net/publication/338511980

[13] A. Patel et al., "Real-time sign language recognition system," ResearchGate, 2025. https://www.researchgate.net/publication/363026081

[14] A. Qureshi et al., "Real-time sign language detection: Empowering the hearing-impaired communication," ScienceDirect, 2024. [Link]

[15] A. Sharma, "Real-time sign language recognition using YOLOv5 for Telugu Sign Language (TSL)," arXiv preprint, 2024. https://arxiv.org/pdf/2406.10231

[16] A. Gupta, "Recognising Words in American Sign Language: A YOLO-based Approach," IJERT, 2025. [Link]

[17] S. Rao, "Sign Language Recognition," IJERT, 2025. https://www.ijert.org/sign-language-recognition-2

[18] N. Mohammed, "An Effective Real-Time Sign Language Recognition using YOLO," SSRN, 2023. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5083947

[19] S. Kumar, "Comparative Analysis of Various YOLO Models for Sign Language Recognition," JISEM Journal, 2025. https://jisem-journal.com/index.php/journal/article/view/8443

[20] A. Wahyudi et al., "Sign language detection and hand gesture recognition systems for real-time interpretation," AIP Conference Proceedings, 2025. https://pubs.aip.org/aip/acp/article/3394/1/040005

[21] Z. Chen et al., "Transfer learning with YOLOv8 for real-time recognition of sign language gestures," ScienceDirect, 2024. [Link]

[22] M. Yu et al., "A survey of deep learning-based sign language recognition: challenges, datasets, and evaluation," CMES, 2024. https://www.techscience.com/CMES/v139n3/55626

[23] R. Singh et al., "Sign language recognition: Static and continuous gesture recognition for Indian sign language," Scientific Reports, 2025. https://www.nature.com/articles/s41598-025-18496-8

[24] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," arXiv, 2015. https://arxiv.org/abs/1506.02640

[25] Wikipedia, "Sign language recognition — definitions, tasks, challenges." https://en.wikipedia.org/wiki/Sign_language_recognition.