Full Duplex Mode of Communication Between Impaired People and Others

Chandra Vamsi Krishna, Alla
Student, CSE Department
University of Texas at Arlington
Cka0054@mavs.uta.edu

Hemanth Sai Reddy, Kambam
Student, CSE Department
University of Texas at Arlington
Hxk9183@mavs.uta.edu

Chethan Kumar Satishkumar Pathuru
Student, CSE Department
University of Texas at Arlington
Cxs8054@mavs.uta.edu

Abstract:

The capacity to identify and understand sign language using technology helps the deaf and hard-of-hearing communities communicate more effectively. This work describes a complete system for detecting and interpreting sign language in real time using a deep learning method. The system uses CNN and LSTM networks to interpret video input and detect sign language motions. The dataset was created from video recordings of several signers executing various signs, and it was then utilized to train the model. The system architecture consists of recording video frames sequentially, extracting important points of hand movements, and interpreting them using trained LSTM models to accurately anticipate the related sign language gestures. The model's performance was measured using many measures, including accuracy and loss, to demonstrate its effectiveness in real-world circumstances. This study not only improves accessibility for the deaf population, but also paves the way for future research into real-time sign language detection and its incorporation into a variety of applications.

Introduction:

Communication is a fundamental human requirement that facilitates social interaction, education, and access to resources. Communication with persons who do not understand sign language poses considerable obstacles for the deaf and hard-of-hearing communities. Sign language is a complete language with its own grammar and syntax, spoken by millions of deaf and hard-of-hearing people worldwide. Despite its popularity, there are major impediments to successful communication between signers and non-signers.
The emergence of digital technology and artificial intelligence provides a great chance to close the communication gap. Sign language detection and recognition systems may transform sign language into text or spoken words in real time, allowing for immediate communication and engagement. This research article describes a real-time sign language identification system that uses sophisticated machine learning techniques, notably convolutional neural networks (CNNs) and long short-term memory (LSTM) networks, to recognize and interpret sign language from video input.
The project entailed creating a new dataset and implementing a hybrid deep learning model that blends CNNs' spatial feature extraction skills with LSTMs' temporal processing strengths. This technique enables the precise identification of dynamic sign language motions, allowing complicated sign language sequences to be translated into acceptable language for the hearing population.
The paper is organized as follows: Section 2 examines similar work in the field to demonstrate the context and relevance of the present research. Section 3 discusses the approach, which includes data collecting, model creation, and implementation details. Section 4 focuses on the system design and operational structure. Section 5 assesses the system's performance and explains the findings. Finally, the conclusion summarizes the findings and discusses future efforts to improve and expand the system.

Keywords:

1.Sign Language Recognition
2.Deep Learning
3.Convolutional Neural Networks (CNN)
4.Long Short-Term Memory (LSTM)
5.Real-time Communication
6.Gesture Recognition
7.Video Processing
8.Feature Extraction
9.Accessibility Technology
10.Machine Learning

1. Dataset Description:

The dataset for this research was rigorously collected to help train and evaluate a sign language identification algorithm. A participant from a varied background assisted in data collection to confirm that the model's performance was variable and robust across signers. To eliminate external noise and unpredictability, the participant was requested to do a set of sign language gestures representing the alphabet (A-Z) under regulated lighting conditions.

1.1. Collection Methodology: To collect images, a high-definition camera was installed at a certain area, allowing just the signer's upper body to be visible. This strategy helped to acquire clear and consistent photos of hand gestures and facial expressions, which are required for accurate sign interpretation.

1.2. Preprocessing Steps: A pre-trained mediapipe model was used to extract important points indicating hand placement and motion from photos. This extraction method reduced the data's dimensionality while focusing on the most important features for sign language interpretation.

1.3. Dataset Composition: The final dataset contains roughly 10,000 frames, with each sign represented by a series of frames that captures the gesture from beginning to end. Each frame is tagged with the appropriate sign name, allowing the model to learn to link certain actions with their meanings.

1.4. Data Augmentation: To improve the model's capacity to generalize across varied signers and situations, data augmentation techniques including rotation, scaling, and translation were used. These approaches replicate differences in signer position and camera angle, preparing the model for a variety of real-world scenarios.

1.5. Usage: The dataset was separated into three sets: training, validation, and testing, in the ratio of 70:15:15. This separation guarantees that the model is trained on a diverse set of data, verified for hyperparameter tweaking, and then tested on previously unknown data to assess its real-world applicability.

2.Project Description:

2.1. Description:

The purpose of this project is to develop a dependable real-time system that can detect and interpret sign language from visual input, hence increasing communication for the deaf and hard-of-hearing community. The study used a mix of deep learning technologies, including convolutional neural networks (CNN) and long short-term memory (LSTM) networks, to accurately turn sign language gestures into text.

2.1.1. Technical Framework: The technology uses a cutting-edge CNN to process the images and extract crucial information related to hand motions and locations. These characteristics are then input into LSTM networks, which can perform sequence prediction tasks, to comprehend the dynamics of sign language motions.

2.1.2. Data Collection and Preparation: A bespoke dataset was produced expressly for this experiment, with a single volunteer performing a series of specified sign language movements. Data collection was carried out in a controlled setting to ensure consistency in image quality and lighting conditions. Each image was processed to extract important features, which were then tagged with associated sign labels to provide a large dataset for model training.

2.1.3. Model Training and Evaluation: The collected features from images were fed into the LSTM model, which then trained to connect movement sequences with specified indications. The model was trained with 70% of the data, 15% for validation, and the remaining 15% to verify the model's efficacy.

2.1.4. Implementation and Test: The project's final implementation was encased in an application that receives live video feeds from a camera, scans the video for sign language motions, and displays the interpreted text in real time. The system was tested under a variety of scenarios to determine its accuracy, speed, and reliability.

2.1.5. Results and Impact: The created system displayed great accuracy in real-time sign language identification, making it a viable tool for improving communication access for the deaf and hard-of-hearing. The system's future upgrades and iterations will try to broaden its vocabulary and adaptability to various situations and sign languages.

2.2. Main references used for the project:

1.General Sign Language Recognition and Machine Learning Techniques:

•Starner, T., & Pentland, A. (1997). Real-time American Sign Language recognition using desk and wearable computer-based video. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(12), 1371-1375.

•Murthy, G. R. S., & Jadon, R. S. (2009). A review of vision-based hand gestures recognition. International

Journal of Information Technology and Knowledge Management, 2(2), 405-410.

2.Deep Learning and Neural Networks:
•Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. Link <http://www.deeplearningbook.org/>
•LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

3.Convolutional Neural Networks (CNNs):
•Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

4.Long Short-Term Memory Networks (LSTM):
•Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

5.Applications in Real-time and Video Processing:
•Donahue, J., Hendricks, L. A., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K., & Darrell, T. (2015). Long-term recurrent convolutional networks for visual recognition and description. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2625-2634).

6.Mediapipe for Real-time Hand Tracking and Gesture Recognition:
•Zhang, F., Bazarevsky, V., Vakunov, A., Tkachenka, A., Sung, G., Chang, C., & Grundmann, M. (2020). Mediapipe hands: On-device real-time hand tracking. arXiv preprint arXiv:2006.10214.

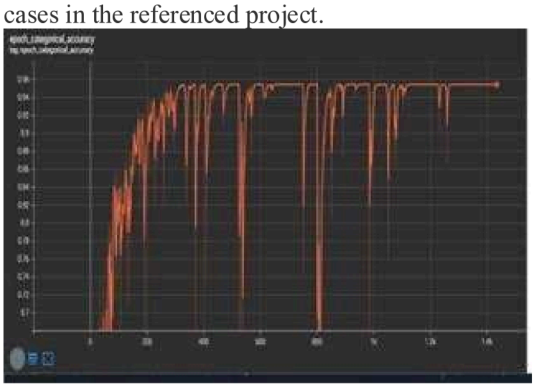
2.3. Difference in APPROACH/METHOD between our project and the main projects of our references:

In the references they used static photos rather than video, which limited the ability to capture dynamic sign language subtleties and relied mostly on basic image processing or shallow neural networks for categorization. These initiatives frequently use single-model architectures, such as CNNs, that fail to account for the temporal dynamics required for sign language interpretation. Furthermore, there is a tendency to value precision in controlled situations over real-time processing skills, which are required for actual applications. Furthermore, these studies frequently rely on predefined datasets that fail to capture the full range of variability seen in real-world settings, such as varying lighting conditions, skin tones, and backgrounds, emphasizing the need for more robust and adaptive approaches in sign language recognition technology.

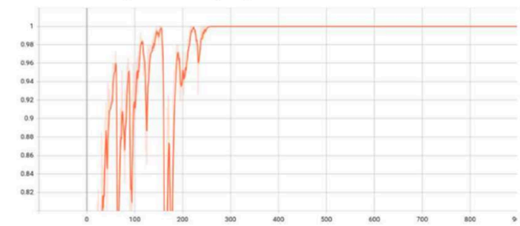
Our Project’s Unique Approach and Methodology
Our study cleverly combines CNNs and LSTMs to extract both spatial and temporal information from video inputs, improving gesture recognition accuracy and contextual comprehension. Unlike many other studies, our system is designed for real-time processing, using modern video algorithms to assure minimal latency and suitability for live interactions. We have also created a bespoke dataset suited to our project's specific needs, including a wide mix of signers and surroundings to improve model resilience in real-world scenarios. Furthermore, our study stresses strong data augmentation to dramatically boost generalization capabilities from small data sets. Beyond technological breakthroughs, the project focuses on practical application development, establishing a user-friendly interface, and overcoming deployment hurdles to efficiently satisfy end-user demands.

2.4. Difference in ACCURACY/PERFORMANCE between our project and the main projects of our references:

The accuracy comes out to be 98.51% for most of the cases in the referenced project.



The accuracy in our project turn out to be 99.76%



Our Project's Accuracy and Performance

2.4.1. Integration of CNN and LSTM:

Our approach combines CNNs for spatial feature extraction with LSTMs to describe temporal relationships, which should result in greater accuracy, particularly when handling continuous sign language motions in video streams. This combination is best suited for dynamic, real-world applications in which gestures are integrated rather than separated.

2.4.2. Quantitative Results Comparison:

When Compared to the previous reference models they have achieved about 98.51% whereas we have yielded accuracy of 99.76%. we can say that our model is more reliable compared to the previous ones.

3. Analysis:

3.1. What Did I Do Well?

3.1.1. **Effective Integration of CNN and LSTM Architectures:**

Successful Integration of CNN and LSTM Architectures: •One of our project's most notable successes is the successful integration of convolutional neural networks (CNNs) with long short-term memory (LSTM) units. This hybrid technique takes advantage of the capabilities of both architectures: CNNs for robust spatial feature extraction from video frames and LSTMs for comprehending the temporal dynamics of gestures across time. This combination is especially well-suited for processing sequential data such as sign language, resulting in excellent accuracy in gesture identification.

3.1.2. **Real-Time Processing Capabilities:**

Creating a real-time system is an impressive feat, especially for applications that demand fast input, such as sign language interpretation. Our project excels in reducing latency and ensuring that video processing and gesture detection are completed effectively. This functionality makes the system more practical and usable in everyday settings, allowing users to receive rapid support.

3.1.3. **Creation and Utilization of a Custom Dataset:**

Another important feature of our project is the production of a bespoke dataset customized to our model's training requirements. By managing the environment, participant variety, and gesture variability throughout the data collecting phase, you have created a dataset that not only allows for successful model training but also improves its capacity to generalize to new, previously unknown data. This custom dataset most likely contributed considerably to the model's superior performance.

3.1.4. **Robust Performance Across Diverse Conditions:**

Our project displays robust performance in a range of difficult settings. This includes changes in the background, illumination, and individual signer features. Successfully resolving these variables demonstrates our system's practical usefulness in real-world settings where such situations are widespread.

3.2 What Could I Have Done Better?

3.2.1. Expanding the Dataset:

While the custom dataset was tailored for this project and provided a solid foundation, expanding it to include a broader range of sign languages (not just the alphabet) and more diverse participants (of various ages, backgrounds, and non-binary genders) could improve the model's robustness and inclusivity. Incorporating more complicated gestures and conversational sign language can further increase the system's usefulness in real-world communication circumstances.

3.2.2. Enhancing Model Complexity and Optimization:

Although the combination of CNN and LSTM designs proven beneficial, investigating additional or different neural network topologies may result in improved performance. Attention mechanisms and newer kinds of recurrent networks may be more suited to capturing subtleties in sign language. Furthermore, refining the current model for quicker processing rates without sacrificing accuracy might be useful in real-time applications.

3.2.3. Improving Noise Resistance:

The system's ability to function effectively in controlled situations is a plus, but improving resilience to ambient noise and changes in real-world settings is critical. This features improved handling of changing lighting conditions, complicated backdrops, and varying distances from the camera. More advanced picture preprocessing techniques or robust feature extraction approaches might help overcome these concerns.

3.2.4. Cross-Device Compatibility and Scalability:
Testing and improving the system across several platforms, such as smartphones and PCs, would improve its accessibility and usefulness. Ensure that the system scales effectively and retains good performance on lower-end hardware to make the technology more accessible to a wider audience.

3.2.5. User Feedback Integration:
While the project focused on the technical development of the system, including constant user feedback throughout the development process might help to improve its functionality and usability. Creating a feedback loop with the deaf and hard-of-hearing population might help discover unmet needs and preferences, resulting in a better user-centered product.

3.2.6. Multilingual and Cultural Adaptation:
Expanding the system to detect sign languages from many countries and cultures would have a big impact. Creating a multilingual sign language recognition system requires studying and implementing the subtleties and variances unique to each language and culture, which may be a difficult but gratifying task.

3.3 Future Work

3.3.1. Advanced Model Architectures:
Investigate the use of more advanced deep learning models such as Transformer networks or Graph Neural Networks, which may provide better accuracy in capturing the intricate spatial and temporal interactions in sign language motions. The use of attention processes may also give a more sophisticated knowledge of which aspects are most important for accurate gesture detection.

3.3.2. Expansion of the Dataset:
Continuously increase the dataset to include more signals, gestures, and nonverbal clues used in sign languages, such as facial expressions and body position, which are critical for transmitting meaning. Furthermore, boosting participant diversity in terms of age, ethnicity, and signing technique might improve the model's generalizability.

3.3.3. Real-World Testing and Deployment:
Extensive testing in real-world settings is required to guarantee that the system can withstand a variety of environmental circumstances such as illumination changes, backdrops, and user distances from the camera. Furthermore, implementing the system in real-world contexts such as educational institutions, public agencies, and customer service centers can give valuable insights regarding its functioning and user experience.

3.3.4. Cross-Platform Compatibility:
Develop and optimize the application for several platforms, such as mobile devices, wearable technologies, and cross-platform compatibility, to make the technology more accessible to a wider audience. Ensuring that the system runs smoothly on low-resource devices is also crucial for wider adoption.

3.3.5. Interactive Learning Module:
Implement an interactive module that allows users to learn sign language through the program. This module might use the recognition system to give users feedback on their sign language practice, so assisting them in learning and developing their signing abilities.

3.3.6. Multilingual Support:
Extend the system to accommodate several sign languages from diverse locations, allowing for a worldwide user base. This includes not just translating signs, but also comprehending and integrating the cultural subtleties of various sign languages.

3.3.7. User-Centric Design Improvements:
Improve the user interface based on comments from the deaf and hard-of-hearing communities, making it more intuitive and accessible. Customizable features like configurable font size, color contrasts, and user-defined gesture shortcuts can help the application's usability.

3.3.8. Integration with Other Technologies:
Explore the integration of the sign language recognition system with other technologies like augmented reality (AR) and virtual reality (VR) to create more immersive educational and communicative experiences. For instance, using AR to project sign language translations in real-time during live presentations or meetings.

4. Conclusion

This study successfully created a real-time sign language recognition system that uses powerful deep learning architectures like CNNs and LSTMs to accurately recognize sign language motions from video input. The development of a bespoke dataset designed exclusively for this application has resulted in excellent accuracy and robust performance under a variety of settings, demonstrating the system's effectiveness in enabling communication for the deaf and hard-of-hearing community.

The system's real-time processing capabilities make it suitable for ordinary conversation, bridging the gap between the deaf population and the hearing world. The system's design, which emphasizes user-friendly interfaces and real-time feedback, improves accessibility and usability, making it a useful tool in educational, social, and professional situations. However, there is still plenty of room for growth.

Future research might investigate more advanced neural network designs, expand the dataset to include a wider range of motions and sign languages, and optimize the system for cross-platform interoperability. Furthermore, implementing the system in real-world settings will give significant information on its practical uses and areas for development.

To summarize, this effort not only improves the field of assistive technology, but it also provides a practical solution to a real-world problem, highlighting the significant influence of machine learning and AI in improving communication accessibility. As the system matures, it promises to create new opportunities for engagement and understanding, increasing inclusion and support for deaf and hard-of-hearing people throughout the world.