

SIGN LANGUAGE DETECTION USING CONVOLUTIONAL NEURAL NETWORK

**Mr. V. Chandra Sekhar Reddy¹, Biroju Ashish², Ballu Premchand³, Digajarla Abhinav⁴,
Modugumpally Manideep⁵**

¹Associate. Professor, CSE Dept, ACE Engineering College, Hyderabad, India.

^{2,3,4}Student, CSE Dept, ACE Engineering College, Hyderabad, India.

DOI: <https://www.doi.org/10.58257/IJPREMS34031>

ABSTRACT

The purpose of Sign Language Detection technology is to use computers to recognize and decipher sign language gestures and movements. This can improve accessibility and communication for the deaf and hard of hearing community. Machine learning algorithms are used to analyze and decipher the signer's hand, arm and body movements through the sign language recognition process. Convolutional neural networks, short term memory networks, and hidden Markov models are popular sign language detection techniques. To detect and decipher sign language gestures, the se algorithms are trained on labeled sign language data using supervised learning techniques. The trained model can there be applied to real time applications, such as sign language translation applications, to convert sign language to text or audio output. In addition to its core function of recognizing and deciphering sign language gestures and movements, Sign language Detection technology also facilities video calling to other individuals.

1. INTRODUCTION

Thank to the technology called Sign Language Detection, computers and other electronic devices can now understand and recognize human gestures. For people who are deaf or hard of hearing and those who cannot speak, technology is designed to improve accessibility and communication. Sign Language Detection is required for proper conversation of sign language gestures and movements into text or speech. To detect and decipher the unique features of sign language, the technology combines computer visions and machine learning techniques. Communications, healthcare and education are as diverse as the fields that could benefit from the use of sign language recognition technology. For example, it can be used to improve communication between healthcare professionals and deaf patients, to provide more accessible internet in formation to the hearing impaired or deaf, and to enable more inclusive communication in the workplace and in public spaces.

2. OBJECTIVES

In our project there are 3 objectives. They can be listed as:

- Enhance Accessibility
- Long Distance Communication
- Time Saving.

3. METHODOLOGY

A labeled dataset of sign language images or videos is collected, containing examples of various sign language gestures. Next, the CNN architecture is designed, typically consisting of multiple convolutional layers followed by pooling layers to extract spatial features from the input data. The network is then trained using supervised learning techniques, where it learns to map input images to corresponding sign language labels. During training, the CNN adjusts its internal parameters to minimize the difference between predicted and actual labels. Once trained, the CNN can be deployed for real-time sign language detection, where it analyzes video streams or images to recognize and interpret sign language gestures accurately. Fine-tuning and optimization of the CNN architecture may be performed to improve its performance in detecting subtle hand movements and gestures, ultimately enhancing the accessibility and communication for the deaf and hard of hearing community.

4. LITERATURE SURVEY

TITLE: DeepASL: Enabling Ubiquitous and Non-Intrusive Word and Sentence-Level Sign Language Translation

AUTHORS: Neidle, Carol et al.

PUBLISHED: ACM Transactions on Accessible Computing, 2019.

ESCRPTION: This paper introduces DeepASL, a system that employs CNNs for word and sentence-level sign language translation. The authors focus on improving the accessibility of sign language translation through non-intrusive methods.

DISADVANTAGES:

Lack of Temporal Information: CNNs are primarily designed for spatial data and may not naturally capture temporal dependencies in dynamic sign language gestures. Additional architectures or post-processing steps might be necessary to address this limitation.

TITLE: Sign Language Recognition using Convolutional Neural Networks for Static Hand Gestures

AUTHORS: Pham, Cuong et al

PUBLISHED: 2017 IEEE-RIVF International Conference on Computing & Communication Technologies - Research, Innovation, and Vision for the Future (RIVF), 2017.

DESCRIPTION: The paper discusses the use of CNNs for recognizing static hand gestures in sign language. It explores the application of deep learning techniques to improve accuracy and robustness.

DISADVANTAGES:

Complexity and Overfitting: Deep CNN architectures can be complex, leading to a risk of overfitting, especially when training data is limited. Regularization techniques and data augmentation may be required to mitigate this issue.

TITLE: Deep Sign: Hybrid CNN-HMM for Continuous American Sign Language Recognition

AUTHORS: Puertas, Enrique Benito et al.

PUBLISHED: Proceedings of the International Joint Conference on Neural Networks (IJCNN), 2017.

DISADVANTAGES:

Complexity: The hybrid CNN-HMM model introduces additional complexity due to the integration of two distinct architectures (CNN and HMM), which might make the system harder to interpret and optimize.

Training Time: Training a hybrid model could require more time compared to a standalone CNN, especially when dealing with a large amount of data or complex architectures.

5. PROPOSED SYSYTEM

The proposed system integrates Sign Language Detection using Convolutional Neural Networks (CNNs) with additional features to enhance accessibility and communication. It includes a comprehensive data pre-processing pipeline for cleaning and augmenting sign language datasets, followed by feature extraction to capture relevant spatial information from video inputs.

The CNN architecture is then trained on the processed data using supervised learning techniques, enabling the network to accurately recognize and interpret sign language gestures in real-time. Additionally, the system incorporates a video calling feature, allowing users to engage in live, visual communication with others. Through rigorous testing and evaluation, the system ensures robust performance and usability, ultimately empowering individuals in the deaf and hard of hearing community to communicate effectively and inclusively.

6. HARDWARE AND SOFTWARE REQUIREMENTS

6.1 HARDWARE REQUIREMENTS:

- Processor: Min. Core i3 processor
- RAM: 2GB (Min.) or 8GB (Recommended)
- Hard Disk Space: 50GB+

6.2 SOFTWARE REQUIREMENTS:

- Programming Language: JavaScript, HTML
- Operating System: Windows 7 or later versions
- of windows.
- Tools: VScode

7. PACKAGES USED

TensorFlow.js

TensorFlow.js allows you to load pre-trained CNN models, perform inference on input data (such as images or video frames), and execute deep learning operations directly in the browser using JavaScript. You can use TensorFlow.js to build and deploy real-time sign language detection applications without the need for server-side processing

OpenCV.js

A JavaScript port of the OpenCV computer vision library, useful for advanced image processing tasks such as edge detection, image segmentation, and feature extraction.

CamanJS

A lightweight image manipulation library for client-side image editing, offering features like filters, adjustments, and layering.

PixiJS

A fast 2D rendering engine that can be used for processing and displaying images efficiently, useful for creating interactive visualizations or effects.

React.js

A popular JavaScript library for building user interfaces, offering a component-based architecture and efficient rendering for creating responsive and interactive web applications.

Vue.js

A progressive JavaScript framework for building UIs, known for its simplicity and flexibility, suitable for developing dynamic and data-driven applications.

AngularJS or Angular:

Comprehensive frameworks for building single-page applications (SPAs), providing tools for data binding, dependency injection, and modular development.

8. TECHNOLOGY DESCRIPTION

JavaScript (JS) is a versatile programming language primarily used for front-end web development. With its lightweight syntax and dynamic nature, JS enables developers to create interactive and dynamic web pages. Its widespread adoption has made it a fundamental skill for web developers.

Node.js, on the other hand, extends JavaScript's capabilities to server-side programming, allowing developers to build scalable and high-performance applications. Leveraging the event-driven, non-blocking I/O model, Node.js excels in handling concurrent requests efficiently. Its vast ecosystem of packages, provided by npm, empowers developers to easily integrate various functionalities into their applications. Together, JavaScript and Node.js form a powerful duo, enabling developers to build full-stack web applications seamlessly, from client to server. Their flexibility and robustness have contributed significantly to the modern web development landscape.

9. CONCLUSION

Improved accessibility and communication for the deaf and hard of hearing communities is on the horizon with the advancement of machine learning-based sign language recognition. This technology, fueled by convolutional neural networks (CNNs), has the potential to accurately interpret sign language gestures and movements in real-time. By integrating such innovations into sign language translation apps with video calling features, individuals facing hearing challenges can communicate seamlessly across various contexts.

While challenges persist in achieving optimal accuracy and reliability, recent strides in research have demonstrated promising avenues. Techniques like long short-term memory networks (LSTMs) and hidden Markov models (HMMs) are also under exploration, contributing to the ongoing refinement of sign language recognition systems. With continued development and collaboration, these advancements hold the promise of significantly enhancing the quality of life and inclusivity for the deaf and hard of hearing populations.

10. FUTURE SCOPE

Advancements in technology

Real-time Translation Apps: Enhancing existing sign language translation apps to provide real-time interpretation of sign language gestures in various contexts, such as video calls, live events, and educational settings.

Improved Accuracy and Robustness: Continued research and development to refine CNN models for better accuracy and robustness in recognizing a wide range of sign language gestures, including nuanced movements and variations across different sign languages.

Multi-modal Integration: Integration of CNN-based sign language detection with other modalities such as voice recognition and natural language processing to create more comprehensive communication systems for deaf and hard of hearing individuals.

Gesture Recognition in Complex Environments: Advancing CNN models to accurately recognize sign language gestures in complex and dynamic environments with varying lighting conditions, backgrounds, and occlusions.

Accessibility in Smart Devices: Incorporating sign language detection capabilities directly into smart devices and wearables, enabling seamless communication for deaf and hard of hearing users in everyday interactions.

Education and Training: Utilizing CNN-based sign language detection for educational purposes, including interactive learning platforms and tools for sign language teachers and learners. **Global Collaboration:** Collaborating with sign language experts and communities worldwide to develop inclusive and culturally sensitive sign language detection systems that cater to diverse linguistic and cultural contexts.

Ethical Considerations and Privacy: Addressing ethical considerations and privacy concerns related to the collection and use of sign language data, ensuring that systems are designed with respect for user autonomy and privacy rights.

11. REFERENCES

- [1] Real time conversion of sign language to speech and prediction of gestures using Artificial Neural Network Author AbeyAbrahamVrohiniProcedia ComputerScience. Volume 143, 2018, Pages 587-594
- [2] Vega, Maria Torres, Decebal Constantin Mocanu, JeroenFamaey, StavrosStavrou, and Antonio Liotta. "Deep Learning for Quality Assessment in Live Video Streaming". IEEE Signal Processing Letters 24, no, 6 (2017): 736-740
- [3] Ken-HaoLiu, Ming-Fang Weng, Chi-Yao Tseng, Yung-Yu Chuang, Member, IEEE, and Ming-Syan Chen, Fellow, "Association and Temporal Rule Mining for Post-Filtering of Semantic Concept Detection in Video", IEEE transactions on multimedia, VOL. 10, NO.2, February 2008.
- [4] Tianchi Huang, Rui-Xiao Zhang, Chao Zhou, and Lifeng Sun, QARC: Video Quality Aware Rate Control for Real-Time Video Streaming based on Deep Reinforcement Learning.arXiv:1805.02482v1 [cs.MM] 7 May 2018.
- [5] M. Jeballi et al. Extension of Hidden Markov Model for Recognizing Large Vocabulary of Sign Language. International Journal of Artificial Intelligence & Applications 4(2); 35-42, 2013