Sign Language Recognition Using Deep Learning: A Review of Methods and Future Directions

1st Shashank Singh Shekhawat
Student AI&DS
Poornima Institute of Engineering & Technology,
Sitapura Jaipur, India
2021pietcashashank051@poornima.org

2nd Anurag Anand Duvey
Ass. Professor
Poornima Institute of Engineering & Technology,
Sitapura Jaipur, India
Anurag.duvey@poornima.org

Abstract

Sign language plays a vital role in facilitating communication for individuals who are hearing-impaired or have speech difficulties. The increasing integration of artificial intelligence and neural networks has greatly accelerated the progress of sign language recognition systems, effectively closing the communication gap between hearing and non-hearing individuals. This paper reviews three key approaches leveraging neural networks: static gesture recognition using convolutional neural networks (cnns), dynamic sequence modeling with long short-term memory (lstm) networks, and real-time gesture processing through mediapipe-integrated architectures. While convolutional neural networks excel at recognizing static gestures, long short-term memory networks demonstrate superior capabilities in handling sequential and dynamic gestures. Each approach is assessed based on its methodology, the data used, its accuracy, and how well it can be applied in real-life situations. The review points out important areas that require further investigation, such as the variety of datasets, the ability to withstand environmental changes, and the efficiency of computational processes, emphasizing the importance of combining different types of data and developing scalable systems. Looking ahead, the focus is on incorporating transformers, lightweight models for edge devices, and seamless integration into wearable and virtual platforms. By overcoming current constraints and embracing cutting-edge technologies, this field has the potential to significantly improve accessibility and promote inclusivity for individuals with hearing impairments.

Introduction

Sign language serves as a crucial means of communication for individuals with hearing and speech impairments, allowing them to convey their thoughts and emotions effectively without relying on spoken words. Unfortunately, the lack of knowledge about sign language among the majority of people poses substantial obstacles to inclusivity and effective communication. To tackle this issue, researchers have delved into the development of automated sign language recognition systems, employing cutting-edge machine learning and neural network technologies.

In the past, sign language recognition was achieved through manual extraction of features and the use of algorithms such as support vector machines (svms) and k-nearest neighbors (knn). These methods frequently encountered challenges with gesture diversity, intricate backgrounds, and the need for immediate processing. In contrast, recent breakthroughs in deep learning have revolutionized the field by empowering models to extract features directly from raw data, enabling them to handle dynamic and intricate inputs.

Convolutional neural networks (cnns) and recurrent neural networks (rnns) have demonstrated significant potential in the field of deep learning. Cnns excel in capturing spatial features, making them ideal for static gesture recognition, while rnns, especially long short-term memory (lstm) networks, are adept at modeling temporal dependencies, making them well-suited for dynamic and continuous signing. This evaluation examines three neural network-based methods for recognizing sign language:

1. A CNN-based system for static sign recognition using GoogleNet.
2. An LSTM-integrated method utilizing Mediapipe for dynamic gesture recognition.
3. A robust LSTM-based framework for sequential gesture processing with high accuracy.

By comparing these methodologies, this paper highlights their strengths, limitations, and areas that can be enhanced. This review seeks to fill existing gaps and suggest future directions, with the goal of enhancing the inclusivity and efficiency of sign language recognition systems.

2. Overview of Research Papers

Recent breakthroughs in neural networks have brought forth innovative methods for recognizing sign language, tackling the difficulties associated with both static and dynamic gesture interpretation. This section provides an overview of three significant papers, discussing their approaches, accomplishments, and potential shortcomings.

Method 1: CNN-Based

The study employed pre-trained GoogleNet CNN for recognizing the ASL alphabet in static conditions, with an accuracy of 91.02%. The approach adopts transfer learning to tune the network for the recognition of 26 alphabets with minimal training of a dataset comprising just 10 images per letter. Data augmentation, including resizing and flipping, enhances robustness. It is computationally light and suitable for isolated alphabet recognition. But, as it uses static gestures for the entire process, it limits the real-world use cases for such applications in a continuous signing scenario. Further, with a small dataset, the model lacks the ability to generalize well with diverse environments and user profiles.

Method 2: Mediapipe+LSTM Based

Mediapipe coupling its keypoint detection with LSTM networks enables reliable real-time dynamic gesture recognition. An ASL gesture requires a 24-frame sequence, normalized using keypoints, thereby reducing computation overload. This system achieves ~91% accuracy in translating dynamic gestures to texts within live communication scenarios. Although this is the case, it is environment-restricted: sensitive to lighting and user positions. In addition, overlapping gestures and rapid changes are some scenarios that present problems; this, in fact, requires improvement to manage realistic coarticulation scenarios.

Method 3: LSTM-Based

This work considers both static and dynamic gestures, using LSTM networks trained on data collected via OpenCV. The model recognizes sequential ASL gestures with an accuracy of 98.13%, easily beating CNN techniques. Full pre-processing of the input, to all intents and purposes, makes the system robust; but it requires the use of confined environments with similar lighting conditions and has a small vocabulary to prevent open conversations. Expanding datasets and considering diverse environments will help to address this problem.

3. Comparison of Methodologies

The section tackles the procedures underpinning each of the three reviewed papers and their respective methodologies, neural architectures, datasets, preprocessing techniques, and performance. While all three works utilize neural networks, from static gesture recognition to dynamic sequence processing, they reveal the breadth of possibilities and limitations of existing sign language recognition systems.

Aspect	CNN-Based	Mediapipe-LSTM Based	LSTM Based
Approach	Static recognition	Dynamic recognition	Sequential recognition
Dataset	10 samples per gesture	24 frames/gesture	OpenCV-based dataset
Accuracy	91.02%	~91%	98.13%
Applications	Educational tools	Live translators	Sign interpreters

4. Strength and Limitation

CNN-Based

Strengths	Limitations
Simple and efficient setup using GoogleNet.	Limited dataset restricts generalizability.
Pre-trained GoogleNet reduces training needs.	No support for real-time or continuous gestures.
High accuracy for static alphabet gestures (91.02%)	

Mediapipe+LSTM Based

Strengths	Limitations
Real-time recognition enabled by Mediapipe preprocessing.	Sensitive to lighting and overlapping gestures.
Low computational demand due to keypoint extraction.	Accuracy (~91%) limited by dataset size and diversity.
Effectively models dynamic gestures using LSTM.	Prefers slower transitions between gestures.

LSTM-Based

Strengths	Limitations
Highest accuracy (98.13%) for dynamic gestures.	Requires controlled environments with consistent lighting.
Comprehensive preprocessing ensures high-quality input.	Limited vocabulary restricts broader applications.
Effective for sequential and dynamic gesture modeling.	High computational demand limits scalability.

5. Future Direction

Further advancements in sign language recognition can address existing limitations and can take leverage from new technologies. These stages of progress include dataset enlargement, robust model design, multi-modal integration, and advanced architecture adoption.

5.1 Dataset Diversification

The characteristics of the dataset across the three papers show important differences in size, diversity, and input modalities. Paper 1 employed a very small dataset consisting of 10 examples per ASL letter; these are limited to static images that can not generalize well. In contrast, Paper 2 captures dynamic gestures in sequences of 24 frames per gesture with hand keypoints and provides a moderate degree of diversity using ASL phrases. Paper 3 uses an OpenCV-based preprocessing approach concerning dynamic gestures and focuses on video sequences, albeit limited to ASL gestures. Future work targets datasets having more than 10,000 samples, multi-language, and multi-modal inputs such as hands, face, and body pose to facilitate generalization and deep learning-based research.

5.2 Strong and Scalable Solution

Robust preprocessing measures such as lighting normalization and lightweight architectures with support on edge devices shall enhance real-world usage.

Aspect	CNN-Based	Mediapipe-LSTM Based	LSTM Based	Future Goal
Accuracy	91.02%	~91%	98.13%	≥98% in real-world conditions
Environmental Robustness	Low	Moderate	Moderate	High (adaptive to lighting, occlusion)
Real-Time Capability	Limited	Yes	Partial	Fully real-time on edge devices

5.3 Adoption of Advanced Modern Architectures

Advanced architectures, such as Vision Transformers (ViT) and Video Transformers (VTN), will provide extra layers of understanding and scalability. In combination with these models and Explainable Artificial Intelligence (XAI), trust and interpretability will significantly improve.

Conclusion

Systems for sign language recognition have come to the fore in their efforts to bridge the communication gaps between hearing and hearing-impaired communities. In addition to showing feature extraction and classification approaches, the studies reviewed highlight the supportive role neural networks-all the way from static recognition using CNNs to dynamic interpretation of gestures with LSTMs-can take in adapting accessibility technologies.

A lot has been done, but a number of challenges still remain open, such as the scourge of very small datasets, semantic sensitivity with respect to atmospheric variations, scalability issues, etc. In the future, efforts will continue to close the gaps with the evolution of diverse datasets, development of robust and adaptive models, and practical integration of these applications.

Emerging technologies, such as transformers, multi-modal learning, and lightweight architectures, promise to provide exciting opportunities to enhance system capabilities. Moreover, partnerships with industries and communities will guarantee that these systems remain inclusive, scalable, and practical. Continued investment in R&D centered on system design will see sign language recognition systems as a key player in fostering inclusion, empowering people, and creating a more accessible world.

References

- Ismail Hakki Yemenoglu, A.F.M. Shahen Shah, Haci Ilhan, *"Deep Convolutional Neural Networks-Based Sign Language Recognition System,"* in IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference, 2021.
- J. R. Pansare, M. Ingle, “Vision-based approach for American Sign Language recognition using Edge Orientation Histogram,” in Proc. International Conference on Image, Vision and Computing (ICIVC), Portsmouth, UK, 2016
- V. N. T. Truong, C. Yang, Q. Tran, “A translator for American sign language to text and speech,” in Proc. IEEE 5th Global Conference on Consumer Electronics, Kyoto, Japan, 2016
- Yegor Matveyas, Assel Mukasheva, Didar Yedilkhan, Arailym Keneskanova, Dastan Kambarov, Dinargul Mukhammejanova, *"Research and Development of Sign Language Recognition System Using Neural Network Algorithm,"* in IEEE 4th International Conference on Smart Information Systems and Technologies (SIST), 2024.
- J. W. Guido, “Learn American Sign Language: Everything You Need to Start Signing”, Wellfleet Press, 2015
- R Rastgoo, K Kiani, S Escalera, “Sign Language Recognition: A Deep Survey”, Expert Systems with Applications Volume 164, February 2021

- Didar Yedilkhan, Assel Mukasheva, “Predictive pricing models to classify potential customers using data-driven approaches”, AIP Conference Proceedings, 18 August 2022
- Soundarya M, Yazhini M, Thirumala sree NS, Sornamalaya NM, Vinitha C, *"Sign Language Recognition Using Machine Learning,"* in International Conference on Advances in Computing, Communication and Applied Informatics, 2024.