International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0512
e ISSN: 2584-2854
Volume: 03
Issue: 11 November 2025
Page No: 3239 - 3241

Bridging the Communication Gap: A Real-Time Static Sign Language Recognition System Using a Convolutional Neural Network

Rithika[1], Maragatham[2]
[1]PG – Computer Science and Engineering (AI & ML), Jansons Institute of Technology (Autonomous), Coimbatore, Tamil Nadu, India
[2]Assistant Professor, Computer Science and Engineering, Jansons Institute of Technology (Autonomous), Coimbatore, Tamil Nadu, India
Email ID: *rithika17032001@gmail.com[1], maragatham@jit.ac.in[2]*

Abstract
Effective communication remains a significant challenge for the global deaf and hard-of-hearing community when interacting with the hearing majority. This paper presents a real-time Sign Language Recognition (SLR) system designed to translate static hand gestures representing alphabetic and numeric characters into text using a standard webcam. The system employs a deep learning approach based on a Convolutional Neural Network (CNN) architecture. A comprehensive pre-processing pipeline including grayscale conversion, Gaussian blurring, and adaptive thresholding ensures robust hand gesture isolation under varying lighting conditions. The implemented CNN model achieved a test accuracy of 98.5% on a dataset of 27 sign classes. The system demonstrates low-latency performance in real-time inference, providing immediate visual feedback by displaying predicted characters and concatenating them into text strings. This work establishes a practical foundation for assistive communication technologies using deep learning.

Keywords: *Convolutional Neural Network; Deep Learning; Real-Time System; Sign Language Recognition; Static Gesture Recognition*

1. Introduction
Human interaction fundamentally depends on effective communication. For millions of deaf and hard-of-hearing individuals worldwide, a significant barrier exists in connecting with the mainly hearing population. This gap often causes social exclusion and makes it difficult to access important areas such as education, healthcare, and jobs. While professional sign language interpreters are crucial, their services are often limited and too expensive. Current technological aids, like sensor-embedded gloves, offer some accuracy but tend to be intrusive and costly. In contrast, many camera-based options struggle to provide the quick, reliable performance needed for smooth, everyday conversations (Zardoshti et al., 2008; Rautaray & Agrawal, 2015).
To address these issues, this research presents a vision-based, real-time Sign Language Recognition (SLR) system. The main goal is to create an accessible software solution that uses a standard webcam to accurately transcribe static manual alphabets (A-Z) and numerical signs (0-9) into text right away. At the core of this approach is a deep learning framework that employs a Convolutional Neural Network (CNN) to enable seamless and natural communication.

1.1. The key contributions of this work are threefold:
- Development of a dedicated image pre-processing workflow for strong hand gesture isolation and feature improvement.
- Creation and testing of a CNN architecture designed for high-accuracy static sign classification.
- Implementation of the trained model within a working, low-latency application that provides real-time visual feedback to the user.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0512
e ISSN: 2584-2854
Volume: 03
Issue: 11 November 2025
Page No: 3239 - 3241

2. Literature Review

The field of SLR has changed greatly. Early systems relied on sensor-based technologies like data gloves (Zardoshti et al., 2008), which were accurate but often intrusive and costly. The move to vision-based methods marked a significant step forward, removing the need for wearable devices. Initial methods used traditional computer vision techniques, such as skin color segmentation to separate the hand from the background, followed by feature extraction using methods like Histogram of Oriented Gradients (HOG). These features were then classified using Support Vector Machines (SVMs) (Singh & Jain, 2012). A key study by Starner et al. (1998) used Hidden Markov Models (HMMs) to recognize a limited range of gestures from video, showing the potential of vision-based systems for real-time use. However, these methods often struggled with changes in lighting, background, and different signers' gestures. The introduction of deep learning, especially the success of Convolutional Neural Networks (CNNs) in image recognition, transformed the SLR landscape (Goodfellow et al., 2016). CNNs automatically learn effective hierarchical features from raw images, eliminating the need for manual feature creation. This was a major advantage, as CNNs can identify crucial details of a hand gesture, which can be hard to design by hand.

Research began to focus on using CNNs for static sign recognition, where each frame shows a single sign. Studies, such as those by Pigou et al. (2015), indicated that CNNs could achieve high accuracy on static sign datasets. The typical architecture includes several convolutional layers for feature extraction, followed by pooling layers for reducing dimensionality, and fully connected layers for classification. Our proposed model follows this established approach.

3. Method

The system architecture has two main phases: Model Training and Real-Time Inference.

3.1. Data Pre-processing Pipeline

The quality of input data is crucial. Our pre-processing pipeline transforms raw images to a format suitable for CNN processing.

- **Grayscale Conversion:** RGB images are converted to grayscale to make the processing easier and focus on shapes.
- **Noise Reduction:** A Gaussian Blur filter smooths the image and removes noise.
- **Hand Segmentation:** We use Adaptive Thresholding (Gaussian method) to create a binary image, effectively isolating the hand from the background under various lighting conditions.
- **Resizing & Normalization:** We resize images to a fixed size of 128x128 pixels. Pixel values are normalized to the range [0, 1] to support model training.

3.2. CNN Model Architecture

The heart of our system is a Sequential CNN model built with Keras/TensorFlow. The architecture is designed for a multi-class classification problem with 27 classes (26 letters plus '0').

1. **Input Layer:** 128 x 128 x 1 (Preprocessed grayscale image)
2. **Feature Extraction:**
    - Conv2D (32 filters, 3x3, ReLU) → MaxPooling2D (2x2)
    - Conv2D (32 filters, 3x3, ReLU) → MaxPooling2D (2x2)
3. **Classification:**
    - Flatten()
    - Dense(128, ReLU) → Dropout(0.4)
    - Dense(96, ReLU) → Dropout(0.4)
    - Dense(64, ReLU)

a. **Output Layer:** Dense(27, Softmax)

The model was compiled using the Adam optimizer and categorical cross-entropy loss function.

3.3. Real-Time Inference System

The trained model runs in a real-time application:

- **Video Capture:** We get the live feed using cv2.VideoCapture(0).
- **Region of Interest (ROI):** A bounding box indicates where the user should position their hand.
- **Frame Processing & Prediction:** Each frame in the ROI goes through the same preprocessing steps. The processed image is fed into the trained model for prediction.
- **Output Display:** The predicted character appears on the screen. A continuous string

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0512
e ISSN: 2584-2854
Volume: 03
Issue: 11 November 2025
Page No: 3239 - 3241

combines recognized characters to form words.

4. Results and Discussion

4.1. Experimental Setup

The dataset included around 2,000 images across 27 classes. We split the data into 80% for training and 20% for testing. The model trained for 20 epochs.

4.2. Quantitative Results

Here is a summary of the model's performance:

- Final Training Accuracy: 99.2% Final Validation Accuracy: 98.8%
- Final Test Accuracy: 98.5%

The training and validation accuracy curves are close, and loss decreased steadily for both sets. This indicates that the model learned well without overfitting, due to the dropout layers.

4.3. Qualitative & Real-Time Performance

We tested the real-time system in a live setting. Key observations included:

- **Low Latency:** Predictions appeared almost instantly after a gesture, meeting the real-time requirement.
- **Robustness:** The system worked well under different indoor lighting, showing the effectiveness of the adaptive thresholding step.
- **Usability:** The interface provided clear and intuitive feedback, displaying both immediate predictions and a running string.

Conclusion

This paper introduced a complete, real-time SLR system for static sign language gestures. The CNN model, supported by a strong pre-processing pipeline, reached a high test accuracy of 98.5% and demonstrated fast performance suitable for practical use. The system serves as a strong proof-of-concept for using deep learning to create accessible assistive technologies. Despite its success, the system has limitations, mainly its focus on static signs. Future work (Phase 2) will aim to develop a combined CNN-LSTM model that can recognize dynamic gestures and continuous sign language sentences, moving toward a more comprehensive translation tool.