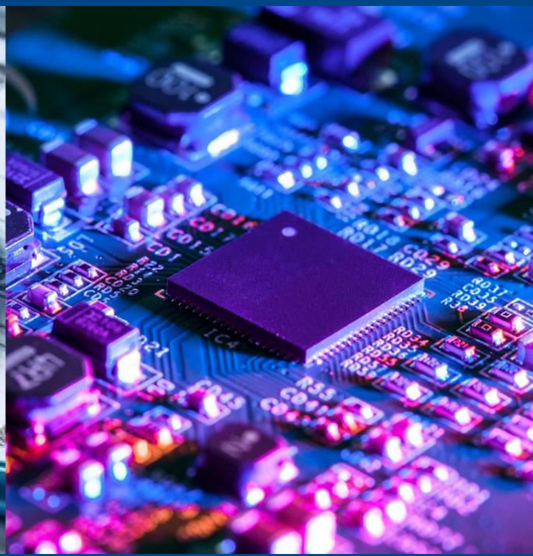
IJMRSET

ISSN: 2582-7219

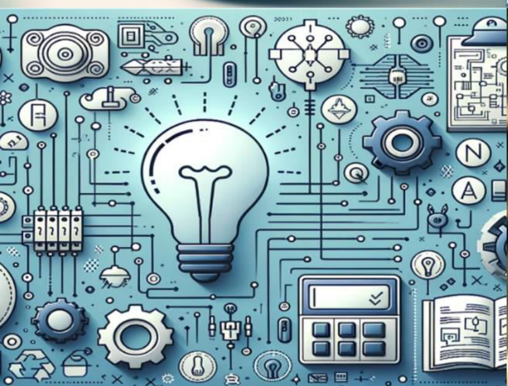




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 4, April 2025

www.ijmrset.com ijmrset@gmail.com +91 6381 907 438 +91 63819 07438

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Real-Time American Sign Language Detector using Computer Vision and Deep Learning

Qureshi Afifa Arif, Dalvi Azka Faiz, Dalvi Haneen Faiz, Prof. Tushar Surwadkar

Department of Artificial Intelligence & Data Science, Rizvi College of Engineering, Mumbai, India

ABSTRACT: This paper presents a real-time American Sign Language (ASL) detector that leverages computer vision and deep learning to bridge communication gaps between sign language users and non-signers. The system uses OpenCV for video capture and preprocessing, and a convolutional neural network (CNN) trained in TensorFlow to classify static hand gestures. The model recognizes 39 distinct ASL gestures, including letters A–Z, digits 0–9, and special tokens such as "space", "del", and "nothing". To ensure stability, a prediction history buffer is used, significantly improving real-time accuracy. The system operates efficiently on standard consumer hardware and shows promise for assistive communication applications.

KEYWORDS: Computer Vision, Sign Language Recognition, Deep Learning, Real-Time Detection, American Sign Language (ASL)

I. INTRODUCTION

Sign language is a vital mode of communication for individuals with hearing or speech impairments. However, its lack of universal understanding poses barriers to inclusive communication. This project addresses this challenge by developing a real-time ASL recognition system that converts hand gestures into textual output using computer vision and machine learning techniques. The primary aim is to create an accessible, efficient, and real-time system that can be operated on consumer-level hardware without requiring specialized sensors.

II. THEORY, METHODOLOGY, AND ALGORITHM

2.1 Theoretical Background

The system is built on the principle of image classification using deep learning. Convolutional Neural Networks (CNNs) are effective in extracting spatial features from images and are widely used in visual recognition tasks. The project also employs OpenCV for real-time video capture and region-of-interest (ROI) processing.

2.2 Methodology

- Model: A CNN trained on a labeled dataset of static ASL gestures.
- Input Handling: Live webcam feed is used for capturing hand gestures.
- Preprocessing: ROI is extracted, resized to 64x64 pixels, normalized, and passed to the model.
- Prediction: The CNN predicts the class label with the highest probability.
- Stabilization: A deque buffer of past predictions ensures smoother output.

2.3 Algorithm Workflow

1. Initialize webcam and load pre-trained model.
2. Capture video frames and define ROI.
3. Preprocess image (resize, normalize, reshape).
4. Predict using CNN and update history buffer.
5. Display most frequent label as the current prediction.
6. Repeat until user exits.

III. LITERATURE REVIEW

Previous research has explored sign language recognition using various methodologies. Early efforts involved sensor-based gloves that translated hand motions into digital signals [1]. However, these systems required specialized hardware and lacked flexibility.

Recent developments have leaned toward computer vision-based approaches using machine learning and deep learning techniques. CNNs, in particular, have been successful in static gesture recognition due to their ability to capture spatial hierarchies in images. Oyedotun and Khashman (2017) applied CNNs to static hand gesture recognition with promising accuracy [2]. Other studies have extended this to continuous sign recognition by integrating temporal models like RNNs or LSTMs [3].

In contrast to these works, this project focuses on a lightweight, real-time system for static ASL detection using only a webcam. It improves on existing models by emphasizing accessibility and live feedback without the need for additional hardware or dynamic gesture recognition.

IV. RESULTS AND DISCUSSION

The system was tested using live webcam input under controlled lighting conditions. It consistently recognized most of the 39 ASL classes with high accuracy when gestures were clearly framed in the region of interest (ROI).

Using a history-based prediction smoothing approach (deque buffer), the system minimized output flickering and maintained stable predictions in real time. This enhancement greatly improved user experience, especially when gestures were held momentarily or transitions occurred between signs.

Some limitations were observed in distinguishing visually similar gestures, such as 'M' and 'N', particularly under poor lighting or partial occlusion. Overall, the model achieved real-time detection speeds (approximately 15–20 frames per second (FPS)) on standard CPU hardware, demonstrating that the design is both efficient and practical.

V. CONCLUSIONS

1. A real-time ASL detection system was successfully implemented using computer vision and CNN-based deep learning techniques.
2. The system accurately detects static ASL gestures and provides live feedback via video overlay.
3. Integration of prediction history improves detection stability without increasing computational cost.
4. The system works on standard hardware without specialized sensors, making it broadly accessible.
5. Limitations include sensitivity to lighting and the absence of support for dynamic sign sequences.
6. Future work could involve mobile deployment, inclusion of dynamic gestures, and multilingual sign support.

REFERENCES

[1] K. Kadous, Machine Recognition of Auslan Signs Using PowerGloves, PhD thesis, UNSW, 2002.
[2] B. Oyedotun and A. Khashman, “Deep learning in vision-based static hand gesture recognition,” Neural Computing and Applications, vol. 28, no. 12, 2017.
[3] N. Camgoz et al., “SubUNets: End-to-end hand shape and continuous sign language recognition,” ICCV, 2017.
[4] T. Sultana et al., “American Sign Language detection using convolutional neural network,” IC4ME2, 2020.

Appendix I – Model Architecture

CNN with convolution, pooling, dense, and softmax layers.
Input size: 64x64 RGB images.
Output: 39 gesture classes.

Appendix II – Dataset Summary

39 classes (A-Z, 0-9, space, del, nothing).
Each class contains several hundred labeled images.
Data augmentation applied during training.

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI:10.15680/IJMRSET.2025.0804216

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Appendix III – Sample Code Snippet

```
def preprocess_frame(frame):
    img = cv2.resize(frame, (64, 64))
    img = cv2.cvtColor(img, cv2.COLOR_BGR2RGB)
    img = img / 255.0
    img = np.expand_dims(img, axis=0)
    return img
```

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com