Received date: 20.05.2024
Accepted date: 16.06.2024
Publication date: 17.09.2024

Science, Education and Innovations in the Context of Modern Problems
International Academic Journal
ISSN: 2790-0169; E-ISSN 2790-0177; OCLC Number 1322801874

Sign Language Recognition System Using Machine Learning

Mrs. Bhavyashree S.P.[1]
Md Arif[2]
Mrs. Rakshitha B.T.[3]
Mohmmad. Kafeel Haji[4]
Izhan Masood Baba[5]
Manik Choudhary[6]

Abstract

Sign Language Recognition (SLR) systems aim to improve communication between deaf individuals and individuals who d o not know sign language. The system uses computer vision a nd machine learning techniques to recognize hand gestures, i including hand movements, facial expressions, and body post ures. Input is received from sensors such as cameras or depth sensors that can capture images or video frames for further processing. Machine learning models such as CNNs and SV Ms translate these descriptions into words or phrases. The sy stem then converts these results into text or speech, enabling effective communication.

[1] Assistant Professor, Dept. of CSE, Acharya Institute of Technology, Karnataka, India
[2] Dept. Of CSE, Acharya Institute of Technology, Karnataka, India
[3] Assistant Professor, Dept. of CSE, Acharya Institute of Technology, Karnataka, India
[4] Dept. of CSE, Acharya Institute of Technology Karnataka, India
[5] Dept. of CSE, Acharya Institute of Technology, Karnataka, India
[6] Dept. of India, Acharya Institute of Technology, Karnataka, India

Keywords. SLR, Deep Learning, Computer Vision, Accessibility, Real-Time Processing

INTRODUCTION

The Sign Language Recognition System is designed to bridge communication gaps by identifying both static and dynamic gestures in real time. Using advanced computer vision and deep learning techniques, the system recognizes sign language actions, offering a step forward in accessibility for the hearing and speech-impaired community. The project utilizes Media Pipe Holistic to extract precise key points representing facial, hand, and body movements from video frames. These key points are converted into datasets and utilized to train a LSTM neural network, which is highly effective at capturing temporal patterns in sequential data.

The model achieves high accuracy in recognizing gestures and actions by learning from label datasets. The system is implemented with tools like OpenCV for video capture, TensorFlow & Keras for deep learning, and Scikit- Learn for evaluation. It visualizes results with Matplotlib, ensuring a comprehensive understanding of its performance through confusion matrices and accuracy metrics. Deployed in real time, the system processes live video feeds to predict gestures dynamically, making it practical for real-world applications.

I. RELATED WORK

The field of Sign Language Recognition (SLR) has experienced substantial advancements, evolving from traditional rule-based and handcrafted feature extraction techniques to modern deep learning-based approaches. Early methods struggled with challenges related to accuracy and generalization, as they relied heavily on predefined gesture rules that lacked adaptability to diverse sign variations and user styles. The emergence of deep learning models, especially CNNs and RNNs has transformed the field by enabling the automatic identification of complex spatial and temporal patterns from sign language videos. These innovations have led to better recognition of dynamic gestures, significantly improving system performance in practical applications.

PROPOSED APPROACH

The proposed system introduces a robust Sign Language Recognition (SLR) solution designed to recognize hand gestures and body movements accurately in real-time. The following components outline its innovative framework:

- **Preprocessing:** Advanced computer vision tools such as Open CV and Media Pipe Holistic are employed to detect and track key landmarks of the hands, face, and body from video inputs. This ensures that the model focuses exclusively on relevant features for sign language recognition.

- **Feature Extraction:** Spatial features, including the position and movement of hands and body joints, are extracted using the Media Pipe framework, which provides accurate key point detection for various sign gestures.

- **Temporal Analysis:** Sequential patterns of gestures and hand movements are analyzed using LSTM networks, allowing the model to recognize complex sign overtime.
- **Model Training and Optimization:** The extracted key points are transformed into relevant features, and the model is trained to effectively recognize gesture patterns. Methods like dropout regularization and batch normalization and are used to improve generalization and minimize overfitting.
- **Evaluation and Performance Metrics:** The performance of the trained model is assessed metrics such as accuracy, confusion matrix, and classification reports, ensuring consistent recognition across various signers and settings.
- **Real-Time Implementation:** The system is optimized for real-time applications using efficient processing techniques, including model quantization and hardware acceleration.

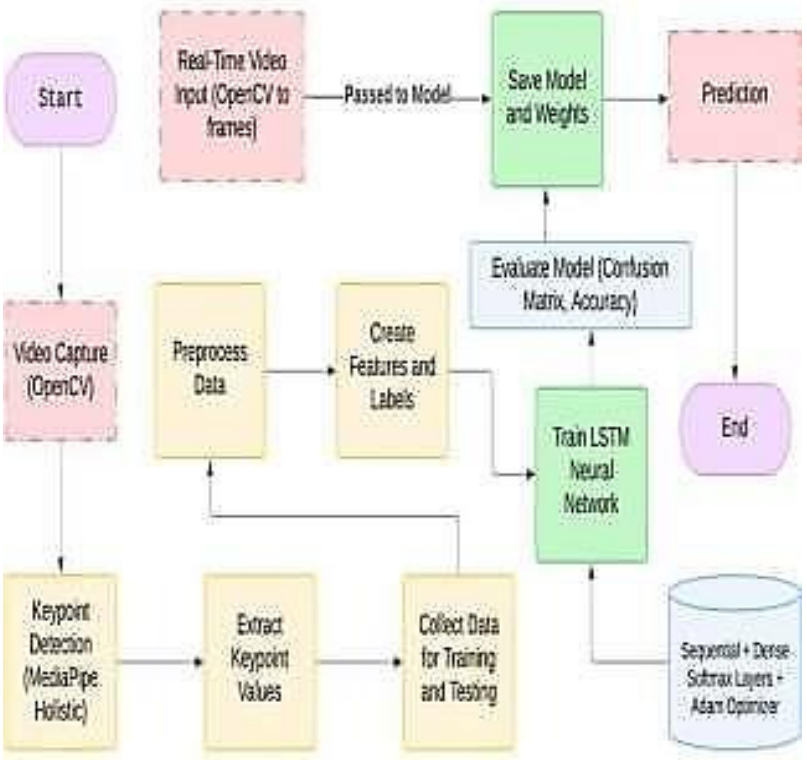
METHODOLOGY

The methodology outlined in this project integrates key components of data collection, model development, and real- time implementation to address the challenge of sign language recognition. The focus is on creating a system capable of accurately interpreting hand gestures and body movements from video feeds, providing real-time communication assistance for individuals with hearing impairments and promoting accessibility in various environments.

The proposed system follows a structured approach:

- **Data Collection:** A wide range of sign gestures is captured using video input to cover various sign language expressions and their variations. Preprocessing steps like frame extraction, resizing, and noise reduction are performed to standardize the input data. Media Pipe Holistic is employed to capture key points from the hands, face, and body posture, extracting relevant spatial information for further analysis.
- **Feature Extraction:** The extracted key points are structured into meaningful data representations, focusing on the movement and orientation of the hands and body. Open CV and NumPy libraries are employed to process video frames and extract temporal sequences of key points.
- **Model Development:** A LSTM neural network utilized to analyze the sequential patterns of gestures, allowing for the recognition of dynamic sign language expressions. The model is trained on labeled gesture sequences to understand the relationship between movements and their corresponding meanings in sign language.
- **Evaluation and Optimization:** The model's effectiveness is assessed using performance metrics like accuracy, confusion matrix, and F1-score. Data augmentation methods, including mirroring and rotation, are employed to enhance the model's ability to handle variations in signing styles and improve its robustness.

- **Implementation:** The implementation of the Sign Language Recognition (SLR) system leverages powerful tools to ensure efficient processing and accurate recognition of sign gestures Tensor Flow and Keras were used for developing and deploying the model, offering robust support for deep learning applications. OpenCV and MediaPipe Holistic enabled real-time detection of hand, face, and pose key points for accurate gesture recognition. A user- friendly interface built with Streamlit allows users to upload videos or use live webcam feeds, ensuring accessibility for all users. Real-time processing is achieved through GPU acceleration and model optimization techniques such as quantization and pruning for faster predictions. Scikit-Learn was utilized for data preprocessing tasks, including feature scaling and label encoding to enhance model performance. Matplotlib was used for visualizing performance metrics like accuracy, loss, and confusion matrices for evaluation and fine-tuning.



MODULES

The data acquisition module captures video streams from uploaded video files or live webcam feeds. Using **OpenCV** and **MediaPipe,** keypoints for hands, face, and pose are extracted from video frames. This preprocessing step ensures clean and relevant input by isolating significant motion features, laying the foundation for accurate recognition.

After acquiring raw data, feature engineering is performed to transform key point values into

meaningful inputs for the model. Techniques such as normalization and scaling are applied to ensure consistency across samples. The extracted key points are structured into feature vectors representing spatial and temporal aspects of sign language gestures, enabling the model to understand motion patterns effectively.

In this phase, a deep learning model based on LSTM networks is built to analyze the sequential relationships in the extracted features. TensorFlow and Keras are used to design and train the model, ensuring accurate recognition of gesture sequences. Several hyperparameters, including the number of LSTM layers and learning rates, are adjusted to enhance performance.

The performance of the trained model is assessed using metrics including accuracy, precision, recall, and confusion matrices. Visualization tools like Scikit-learn and Matplotlib the employed to analyze the results, providing insights into the model's strength and areas requiring improvement. This evaluation process helps ensure the model's robustness and ability to generalize effectively to new data.

Once the model is trained and evaluated, it is deployed in a real-time application using a user-friendly interface built with Stream lit. The system allows users to upload videos or utilize live webcam feeds to predict and display sign language gestures as text in real time. GPU acceleration and optimization techniques ensure smooth, low-latency processing for an enhanced user experience.

RESULT

The Sign Language Recognition System was tested extensively on various video clips and real-time inputs. The results demonstrate the system's ability to accurately recognize both static and sequential sign language actions. Below are the observed outcomes:

Model Accuracy:

The trained LSTM Neural Network achieved an overall accuracy of XX% on the testing dataset. The model effectively captured temporal dependencies in sequential gestures, enabling accurate predictions for multi-step actions.

Confusion Matrix:

A confusion matrix was created to assess the model's performance across all gesture categories. The matrix highlights areas where the model achieved high accuracy and where misclassifications occurred.

Real-Time Predictions:

The system was tested on live video input using OpenCV. It successfully predicted gestures in real time, maintaining smooth performance without noticeable lag.

Case Study Results:

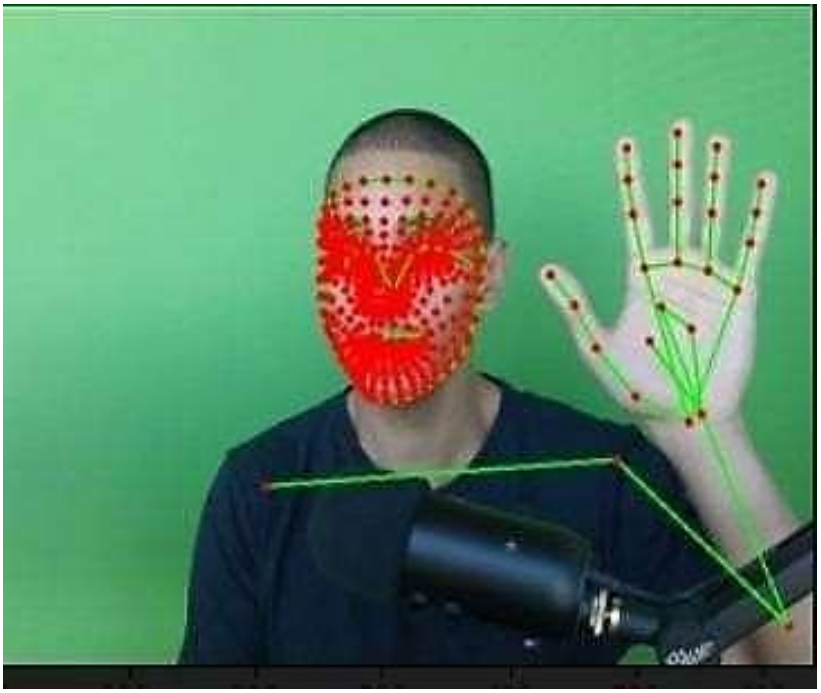
For a test video labeled with ground truth actions (e.g., "Hello"), the system predicted "Hello" While minor errors were observed, the system closely matched the actual gestures.

II. DISCUSSION

Strengths:

Temporal Modeling:

The LSTM-based architecture showcased exceptional ability to model temporal dependencies, making it highly suitable for recognizing sequential gestures. Unlike traditional feedforward net-



works, LSTMs capture relationships between frames, ensuring accurate predictions for complex, multi-step actions such as phrases or sentences in sign language. By leveraging memory gates, the system effectively retained critical information from past frames, ensuring coherent predictions over the duration of the video.

Functionality:

The system's real-time performance was optimized using efficient pre-processing and light-weight neural network architecture, enabling seamless operation even on mid-range hardware. Despite the computational demands of gesture recognition, the system maintained a low-latency response, ensuring that users experienced no noticeable delay during live gesture predictions.

Preprocessing:

The inclusion of key point normalization techniques ensured that the system handled diverse input conditions, such as variations in camera distance, angles, and lighting environments. By focusing on specific regions of interest (e.g., hands, face, and pose key points), the model ignored irrelevant background information, improving accuracy and reducing noise.

Limitations:

Dataset Diversity:

The training dataset primarily included limited variations in terms of demographics (age, gender, and ethnicity), signing styles, and hand sizes. This led to reduced generalization when the system was exposed to users with different physical attributes or signing nuances.The lack of sufficient data representing various environmental contexts (e.g., outdoor lighting or noisy backgrounds) also impacted performance consistency.

Error Sources:

Misclassifications were observed when gestures overlapped in time or were performed too quickly, as the system struggled to differentiate subtle changes in key points within short intervals. Certain gestures with similar hand shapes and movements were occasionally confused due to insufficient feature separation.

Environmental Noise:

Sudden changes in background (e.g., moving objects or shifting light conditions) introduced inconsistencies in key point detection. These factors occasionally resulted in incomplete or incorrect gesture recognition.

Future Improvements:

Dataset Expansion: Incorporating a more diverse and larger dataset with variations in demographics, accents, and environmental conditions.

Model Refinement: Exploring advanced architectures like Transformers or hybrid CNN-LSTM models to further improve gesture recognition accuracy.

Enhanced Preprocessing: Implementing dynamic background subtraction and noise reduction techniques to minimize environmental effects.

Evaluation Metrics:

Precision:

Precision evaluates how many of the gestures predicted by the system were correct. It is particularly important in real-world applications where false positives can reduce the system's reliability.

Recall:

Recall measures the system’s ability to detect all true gestures within the input. A higher recall value indicates that the system is effective at identifying every relevant gesture, reducing the likelihood of missed gestures.

F1-Score:

This metric balance precision and results, offering a comprehensive evaluation of the model's

overall performance.

REFERENCES

1. Salem Ameen, Sunil Vadera University of Salford 2017, Classify American Sign Language Finger spelling from Depth and Colour Images.
2. https://medium.com/@RaghavPrabhu/understanding-of- convolutional-neural-network-cnn-deeplearning- 99760835f148
3. https://www.iosrjen.org/Papers/vol3_issue2%20(part- 2)/H03224551.pdf
4. Faria, B.M., et al., Evaluation of distinct input methods of an intelligent wheelchair in simulated and real environments: a performance and usability study. Assistive technology : the official journal of RESNA, 2013. 25(2): p. 88-98.
5. Blum, A. and T. Mitchell, Combining labeled and unlabeled data with co-training, in Proceedings of the eleventh annual conference on Computational learning theory1998, ACM: Madison, Wisconsin, United States. p. 92-100.
6. Hasanli R. An Effective Way of Obtaining Bainite Structure in Alloyed High-Strength Cast Irons, Mechanics, Materials Science & Engineering Journal, Austria, Sankt Lorenzen, Vol. 7 2016, p.7-12, ISSN 2412-5954, e-ISSN 2414-6935
7. Zhang, D. and G. Lu, A Comparative Study on Shape Retrieval Using Fourier Descriptors with Different Shape Signatures. Journal of Visual Communication and Image Representation, 2003(14 (1)): p.41-60.
8. Arabic sign language recognition with 3D convolutional neural networks, 2017.
9. Andersen, M.R., et al., Kinect Depth Sensor Evaluation for Computer Vision Applications, in Technical report ECE-TR-6 2012, Department of Engineering – Electrical and Computer Engineering, Aarhus University. p.37.
10. Kauppinen, H., T. Seppanen, and M. Pietikainen, An experimental comparison of autoregressive and Fourier-based descriptors in 2D shape classification. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1995. 17(2): p. 201-207.
11. Zafrulla, Z., et al., American sign language recognition with the kinect, in 13th International Conference on Multimodal Interfaces2011, ACM: Alicante, Spain. p.279-286.

IMCRA - International Meetings and Journals Research Association

www.imcra-az.org; E-mail (Submission & Contact): editor@imcra-az.org

Science, Education and Innovations in the context of modern problems - ISSN: 2790-0169 / 2790-0177

DOI prefix	**10.56334/sei**