International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

A Review on Sign Language Recognition and Learning System

Akhilesh K B[1], Anjana T A[2], Rubeena Abdul Jabbar[3], Shifaana M S[4], Gishma K M[5]

[1,2,3,4]*Student, Department of computer science and engineering, Universal Engineering College, Kerala, India.*
[5] *Assistant Professor, Department of computer science and engineering, Universal Engineering College, Kerala, India.*

---***---

Abstract - This paper presents innovative approaches to various aspects of sign language recognition and learning, catering to the diverse needs within this field. It introduces a system utilizing 2D image sampling and concatenation, trained with convolutional neural networks, achieving accurate and robust sign recognition even with low-cost cameras. Another system, SignQuiz, offers a cost-effective web-based solution for learning finger-spelled signs in Indian Sign Language (ISL), outperforming traditional printed mediums. Additionally, a dynamic hand gesture recognition system employing deep learning architectures demonstrates superior performance over existing methods, addressing challenges such as hand segmentation and sequence modeling. Furthermore, an efficient deep convolutional neural network approach for hand gesture recognition, including transfer learning, achieves high recognition rates across various datasets. Finally, a robust model for static sign recognition in ISL, utilizing CNNs and evaluating on a diverse dataset, attains exceptional accuracy and outperforms previous works. These contributions collectively advance the field of sign language recognition and learning, offering solutions that are accurate, cost-effective, and efficient, thereby facilitating better communication and interaction for the hearing-impaired community.

Key Words: CNN, Sign language recognition, Hand gesture recognition, Machine learning, Transfer Learning, SVM.

1.INTRODUCTION

There are approximately 466 million people worldwide who have hearing loss, including deaf people, of which 34 million are children. For the deaf community, sign language is an essential form of communication. Several sign languages are used worldwide, such as Indian Sign Language (ISL). But there are still barriers to education and work for the deaf, especially in developing nations like India. The World Health Organization promotes sign language instruction and formal recognition in order to improve accessibility and inclusivity, acknowledging the significance of sign language.

A key component of human-computer interaction, hand gesture detection is used in many industries, including video games, smart TV control, and sign language interpretation. Because of the subtleties of hand and body movements, sign language presents particular difficulties for recognition as a sophisticated system of manual gestures. Due to the growing number of deaf people and the widespread use of touchless devices, automatic hand gesture detection has become more popular.

Numerous methods have been used in the recognition of sign language, such as convolutional neural networks (CNNs), hand trajectory extraction, and RGB-D cameras. Nonetheless, obstacles continue to exist, such as the requirement for economical and effective techniques to manage the substantial datasets in jobs related to sign language recognition. Notwithstanding these obstacles, sign language is still essential for promoting communication and closing the gap between the hearing and the deaf communities, highlighting the need for ongoing study and advancement in this area.

2. RELATED WORK

Jayesh Gangrade and Jyoti Bharti [1] explores the significance of hand gestures as a natural mode of human-computer interaction and the growing interest in computer vision-based systems, particularly convolutional neural networks (CNNs), for hand gesture recognition. Emphasizing the potential of CNNs in transforming human-technology interaction, the paper discusses challenges such as variations in posture and lighting conditions. It presents a comprehensive survey of the stateof-the-art in vision-based hand gesture recognition using CNNs, highlighting their applications in gaming, human-computer interaction, sign language recognition, and virtual reality. The focus shifts to the specific context of Indian Sign Language (ISL), proposing a CNN-based system achieving a remarkable 93.5% recognition accuracy. The study underscores the effectiveness of CNNs in ISL recognition while acknowledging the importance of diverse datasets for training. It discusses the choice of methods for hand gesture recognition, noting the potential of CNNs and the relevance of other methods like SVM. Future scopes include enhancing recognition accuracy, exploring advanced CNN architectures, and addressing real-time implementation challenges, emphasizing the need for large and diverse datasets for robust system performance.

Shagun Katoch, Varsha Singh, Uma Shanker Tiwary [2] presented a comprehensive methodology for recognizing Indian Sign Language (ISL) gestures. The authors commence with the crucial step of data collection, assembling a dataset essential for training and evaluating the proposed recognition system. Leveraging the Speeded-Up Robust Features (SURF) algorithm, known for its resilience to scale and orientation variations, the authors extract distinctive features from the sign language gestures. Subsequently, these SURF features are utilized to train a Support Vector Machine (SVM), a machine learning algorithm adept at classifying data points into different categories, likely corresponding to specific ISL gestures. Concurrently, Convolutional Neural Networks (CNN) are

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

employed for sign language recognition, capitalizing on their effectiveness in image-related tasks. The paper advocates a comparison and fusion of outputs from the SVM and CNN models to enhance overall recognition accuracy, a strategy often employed to synergize the strengths of multiple models. The proposed system's performance is rigorously evaluated using metrics such as accuracy, precision, recall, and F1 score, providing comprehensive insights into its effectiveness in recognizing Indian Sign Language gestures. The integration of traditional computer vision techniques, represented by SURF, with state-of-the-art deep learning methods, as embodied by SVM and CNN, showcases a robust and multifaceted approach towards achieving accurate and effective ISL recognition.

Yulius Obia, Kent Samuel Claudioa, Vetri Marvel Budimana, Said Achmada, Aditya Kurniawan [3] addresses the challenge of communication faced by deaf individuals by developing a desktop application that recognizes American Sign Language (ASL) gestures in real-time and converts them into text. Rooted in the fields of Artificial Intelligence (AI) and Computer Vision, the application employs Convolutional Neural Networks (CNN) for gesture classification, utilizing a dataset from Kaggle. The two-layer CNN model undergoes training and testing phases, with parameters such as dataset size and iteration numbers affecting accuracy. The application's architecture involves image processing, symbol prediction, and auto correct features using the Hunspell library. The Sign Language To Text Converter application, developed in Python using TensorFlow, NumPy, OpenCV, and other libraries, achieved an impressive 96.3% accuracy for the 26 letters of the alphabet. The GUI incorporates an ASL symbols reference for user convenience. This innovative solution, offering real-time sign language interpretation, addresses the communication gap for special needs individuals, providing a reliable means of interaction. The comprehensive methodology, dataset analysis, and model modifications contribute to the success of the application, making it a promising tool for enhancing inclusivity and accessibility in various fields.

Arun Singh,Ankita Wadhawan, Manik Rakhra, Usha Mittal, Ahmed Al Ahdal, Shambhu Kumar Jha [4] focuses on addressing communication challenges faced by the Deaf and Hard of Hearing population in India through the development of an Indian Sign Language Recognition System (ISLRS). Recognizing the limitations of conventional communication methods for the hearing-impaired, the system aims to facilitate interaction in public settings without the need for interpreters. Leveraging Computer Vision, Convolutional Neural Networks (CNN), Long ShortTerm Memory (LSTM), and Gated Recurrent Unit (GRU), the model is trained and tested on dynamic sign video clips, achieving a training accuracy of 70%. The proposed solution employs gesture detection techniques and preprocessing using Scale Invariant Feature Transform, Shape Descriptors, and Histogram of Oriented Gradients algorithms. The output, representing Indian Sign Language numerals, alphabets, and common words, is converted into both text and speech. The system not only serves as an alternative and effective means of communication for the speech, hearing, and visually impaired but also lays the groundwork for future research to enhance model accuracy and expand the vocabulary.

Ahmed Kasapbasi, Ahmed Eltayeb Ahmed Elbushra, Omar Al-Hardanee, Arif Yilmaz [5] addresses the global issue of hearing impairment affecting over 5% of the population by focusing on the development of a robust sign language recognition system, specifically targeting the American Sign Language alphabet (ASLA). The research emphasizes the challenges faced in recognizing hand poses and gestures, with a particular focus on the complexities associated with similar-looking signs. The paper introduces a novel ASLA dataset captured through laptop and smartphone cameras, catering to variations in lighting and distance. Leveraging Convolutional Neural Network (CNN) architecture, the proposed deep learning model demonstrates exceptional accuracy of 99.38%, outperforming existing datasets. The CNN design, featuring three convolutional layers, showcases optimized speed and accuracy in interpreting sign language gestures. By incorporating three datasets, including an in-house creation, the study contributes significantly to the field of machine learning and deep learning for sign language recognition systems, providing a promising solution with superior accuracy that can benefit hearing-impaired individuals in real-world applications.

3. DESIGN AND ANALYSIS

By utilizing the capabilities of cutting-edge neural networks, the deep learning-based sign language recognition system for static signs seeks to transform communication accessibility for people with hearing impairments. The system's convolutional neural network (CNN) architecture allows it to recognize and understand static signs with accuracy, making it possible to convert sign language movements into textual or audio outputs with ease. This creative solution promotes equitable participation and understanding for people with a range of communication needs while simultaneously addressing the shortcomings of conventional sign language recognition technologies and improving the general inclusivity of communication platforms. Deep learning techniques are integrated into the system to ensure robust performance and flexibility. This makes the system a promising advancement in promoting inclusion and increasing accessibility for the community of hearing-impaired individuals.Data collection, image preprocessing, CNN classifier training, and testing are the four main stages of the sign language recognition system. The data flow diagram that shows the system's functioning model is described in Figure 3.1. The first stage is called data collection, and it involves using a camera to gather RGB data from static signs. Next, image resizing and normalization are applied as preprocessing steps to the gathered sign images. For later use, these normalized photos are kept in the data store. The suggested system is trained using a CNN classifier in the next phase, and testing is conducted using the trained model. The testing process, which comes last, involves fine-tuning the CNN design parameters until the results achieve the required accuracy.[8:23 PM, 3/30/2024] Shifaana Ms: CNN uses convolutions to learn features with higher order that are present in the data. The CNN architecture performs admirably when it comes to object detection, including picture recognition. They are able to identify faces, people, street signs, and other aspects of visual information. Many CNN versions exist, but they are all predicated on the layer pattern that is present, as Fig. 3.2 illustrates. The various parts of the CNN architecture include several layer and activation function kinds. The listing explains the function and goal of a few widely used layers, which are covered in more detail below. Layer of convolution The convolutional layer is one of CNN architecture's fundamental building pieces. Convolutional layers employ a patch to alter the input data.The listing explains the function and goal of a

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

few widely used layers, which are covered in more detail below. Layer of convolution The convolutional layer is one of CNN architecture's fundamental building pieces. A patch of locally linked neurons from the preceding layer is used by convolutional layers to alter the input data. The layer that sits between the input layer's area of neurons and the output layer's locally connected weights will compute the dot product together. The mathematical process that characterizes the rule for combining two sets of data is called a convolution. After applying a convolution filter or kernel to an input, the convolution operation outputs a feature map.The listing explains the function and goal of a few widely used layers, which are covered in more detail below. Layer of convolution The convolutional layer is one of CNN architecture's fundamental building pieces. A patch of locally linked neurons from the preceding layer is used by convolutional layers to alter the input data. The layer that sits between the input layer's area of neurons and the output layer's locally connected weights will compute the dot product together. The mathematical process that characterizes the rule for combining two sets of data is called a convolution. After applying a convolution filter or kernel to an input, the convolution operation outputs a feature map.

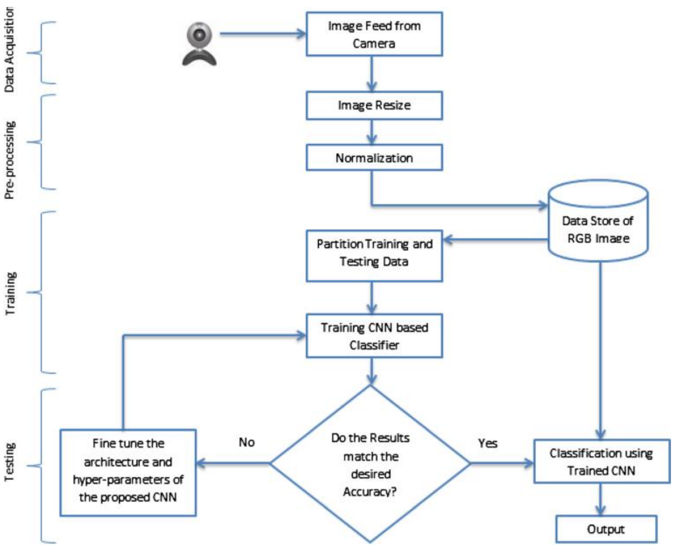

Fig -3.1: System flowchart.

The mathematical process that characterizes the rule for combining two sets of data is called a convolution. As seen in Fig. 3.3, the convolution operation takes an input, applies a convolution filter or kernel, and outputs a feature map. This process shows how the kernel slides over the input data to create the complex output data. Every step creates a single value in the output feature map by multiplying the values of the input data by the kernel within its bounds. Data collection, image preprocessing, CNN classifier training, and testing are the four main stages of the sign language recognition system. The data flow diagram that shows the system's functioning model is described in Figure 3.3. The first stage is called data collection, and it involves using a camera to gather RGB data from static signs. Next, image resizing and normalization are applied as preprocessing steps to the gathered sign images. For later use, these normalized photos are kept in the data store. The system is trained using a CNN classifier in the next phase, and testing is subsequently conducted using the trained model. The testing process, which comes last, involves fine-tuning the CNN design parameters until the results achieve the required accuracy.

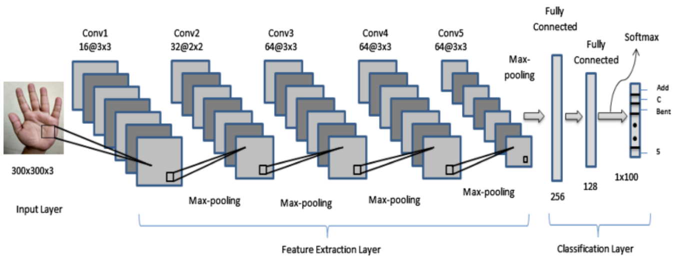

Fig-3.2: High-level general CNN architecture.

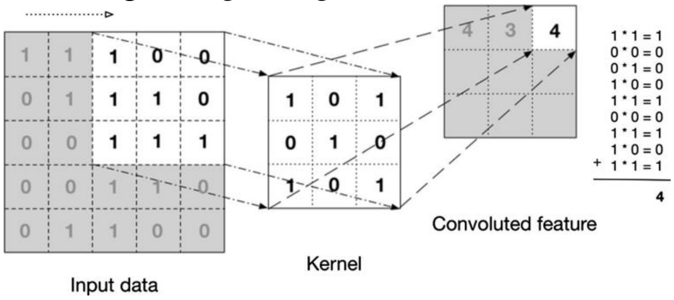

Fig-3.3: The convolutional operation..

4. DISCUSSION

The experiment detailed in the provided text presents a comprehensive analysis of an approach to sign language gesture recognition using deep learning techniques, focusing on addressing challenges such as overfitting, parameter tuning, and model architecture. Several methods were employed to improve the performance of the recognition system, including dropout, L2 regularization, weight initialization techniques, and optimization of hyperparameters like batch size and learning rate.Overfitting, a common issue in deep learning models, was effectively mitigated using dropout regularization, which randomly sets a fraction of the weights to zero during training, preventing the network from relying too much on specific features. The optimal dropout rates were determined through experimentation, with values of 0.8 for convolutional layers and 0.45 for fully connected layers. Additionally, L2 regularization was employed to further prevent overfitting, with an optimal regularization parameter (λ) of 0.0001 identified through experimentation.The choice of weight initialization method also played a crucial role in the performance of the model. Stochastic initialization with a normal distribution and Xavier initialization were compared, with Xavier initialization demonstrating faster convergence and improved performance.

Furthermore, hyperparameters such as batch size were carefully selected through experimentation, with a batch size of 25 identified as optimal. This choice likely facilitated more frequent updates to model weights, leading to better generalization.The experimental results show significant improvements in the final test success rate, reaching up to 86% accuracy. Evaluation metrics such as precision, recall, and F1-score were computed to assess the performance of the Indian Sign Language recognition system comprehensively.The study also compared different model architectures, including MLP and LSTM, with MLP achieving better recognition accuracy in all scenarios. Additionally, a comparison with state-of-the-art methods highlighted the effectiveness of the proposed approach, particularly in terms of accuracy and computational efficiency.Despite the advancements achieved, challenges such as variations in signer gestures, motion blur, and lighting conditions were acknowledged as areas for further improvement. Overall, the study provides valuable insights into the design and optimization of deep learning models for sign language gesture recognition, with implications for real-world applications such as assistive technology and human-computer interaction.

5. CONCLUSIONS

The research discussed in this paper highlights significant progress and promising directions in the realm of sign language learning and recognition systems. Leveraging Deep Neural Networks (DNNs), innovative web-based applications, provide customizable and accessible learning experiences, benefiting both deaf and non-deaf individuals and promoting inclusivity within the deaf community. Moreover, the development of novel sign language learning systems based on 2D image sampling and convolutional neural networks demonstrates impressive accuracy and efficiency, even with limited resources, addressing challenges faced by conventional sign recognition methods, especially in resource-constrained settings. Additionally, the proposed dynamic hand gesture recognition systems, integrating multiple deep learning techniques, showcase superior performance in recognizing complex sign language gestures, offering potential applications in communication aids and interactive human-computer interfaces. Overall, these findings contribute to advancing the field and hold promise for enhancing accessibility, inclusivity, and communication for the global deaf and hearing-impaired populations, with continued research expected to further refine and improve sign language recognition systems.

ACKNOWLEDGEMENT

The authors extend sincere gratitude to Mrs. Gishma K M for her Guidance and project coordinators, Mrs. Najla Nazar and Mrs. Gishma K M for mentorship. Their expertise played a vital role in shaping the direction and focus of this research.

REFERENCES

1. Jayesh Gangrade and Jyoti Bharti,”Vision based hand gesture recognition for Indian sign language using convolution neural network”,2020.
2. Shagun Katoch, Varsha Singh, Uma Shanker Tiwary,”Indian sign language recognition system using SURF with SVM and CNN”,2022.
3. Yulius Obia, Kent Samuel Claudioa, Vetri Marvel Budimana, Said Achmada, Aditya Kurniawan, ”Sign language recognition system for communicating to people with disabilities”,2022.
4. Arun Singh,Ankita Wadhawan, Manik Rakhra, Usha Mittal, Ahmed Al Ahdal, Shambhu Kumar Jha, ”Indian Sign Language Recognition System for Dynamic Signs”, 2022.
5. Ahmed Kasapbasi, Ahmed Eltayeb Ahmed Elbushra, Omar Al-Hardanee, Arif Yilmaz, ”DeepASLR: A CNN based human computer interface for American Sign Language recognition for hearing-impaired individuals”,2022.