Bridging Gaps For The Deaf With Real-Time Regional Sign Language Conversion

Aravind Haridas, Aswin K, Harinandanan B S, Vishnu S Sekhar
Department of Computer Science and Engineering, Amrita School of Computing
Amrita Vishwa Vidyapeetham, Amritapuri, Kerala, India

***Abstract*—This paper presents a machine learning system for American Sign Language (ASL) recognition using Convolutional Neural Networks (CNNs). Through extensive training on a curated dataset, the system achieves accurate interpretation of ASL gestures, addressing communication barriers for individuals with hearing impairments.**

***Index Terms*—sign language, gesture recognition, CNN**

I. Introduction

Sign language is a unique form of communication used by individuals who are deaf or mute. It involves using hand movements, gestures, facial expressions, and body language to convey meaning. Since spoken language isn't accessible to deaf individuals, sign language serves as their primary mode of communication. However, understanding sign language poses a challenge for those who do not know it, creating a communication barrier between deaf and hearing individuals. This research aims to address this challenge by developing a computer-based system that can recognize and interpret sign language gestures. By enabling computers to understand sign language, we can facilitate better communication and inclusivity for all.

Fig. 1. Signs

II. Related Works

Significant research has been conducted to improve sign language recognition and bridge communication gaps. This section explores various approaches, including utilizing hand-crafted features for word-level recognition, applying machine learning techniques, and leveraging deep learning architectures. We will also examine research focused on real-time applications for resource-constrained devices and studies exploring complex architectures to achieve higher accuracy.

Deep Learning Approaches: Deep learning has emerged as a powerful tool in this field. A Survey on Deep Learning Techniques for Sign Language Recognition [1] provides a comprehensive overview of various deep learning approaches. Furthermore, research has explored leveraging existing object detection frameworks like TensorFlow for sign language tasks [6].

Real-time Applications and Resource-Constrained Devices: Recent research focuses on real-time applications and resource-constrained devices. An example of this focus is "Real-time Continuous Sign Language Recognition on Resource-Constrained Devices" [IEEE Transactions on Circuits and Systems for Video Technology, 2018], which addresses real-time recognition on limited-resource devices.

Complex Architectures for Improved Accuracy: To achieve higher accuracy, some studies delve into more complex architectures. This includes research on multi-stream deep learning architectures [3] and end-to-end sign language recognition with hierarchical attention mechanisms [4]. Additionally, research has investigated the use of 3D convolutional neural networks for sign language recognition tasks [5].

III. Methodology and Algorithms

In our study, we developed a systematic approach to enable computers to understand sign language efficiently. Firstly, we curated a large dataset comprising thousands of images depicting various sign language gestures. These images underwent meticulous preprocessing to enhance clarity and consistency, facilitating effective learning by the computer. Utilizing Convolutional Neural Networks (CNNs), advanced algorithms adept at recognizing patterns in visual data, we trained the computer to interpret sign language gestures. This training process involved presenting the CNN with the images of hand signs and teaching it to associate each gesture with its corresponding meaning. Similar to teaching a child to recognize objects, the CNN gradually learned to interpret the hand signs depicted in the images. Through iterative training sessions, the CNN refined its understanding, achieving

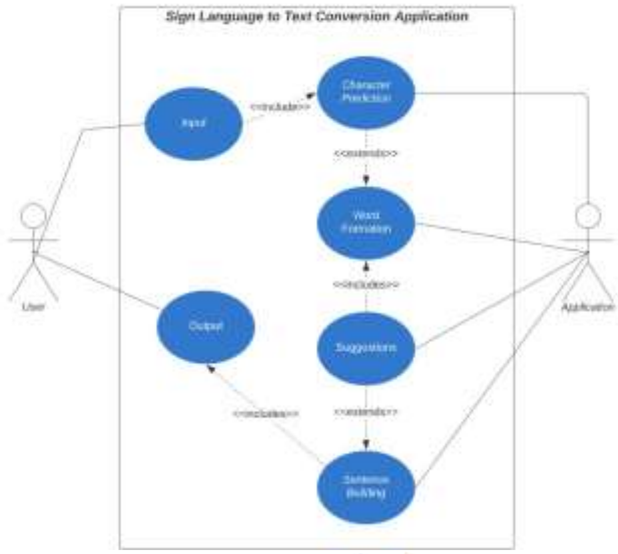

Fig. 2. System Design

a high level of accuracy in identifying and categorizing sign language gestures. Our methodology emphasized not only data acquisition but also preprocessing and model training, ensuring the computer's proficiency in comprehending and responding to sign language cues. This systematic approach laid the groundwork for the successful implementation of a robust sign language recognition system, promising improved communication accessibility for individuals with hearing impairments.

A. Data Collection

To develop our sign language recognition system, we gathered a dataset of sign language gestures. We utilized webcams to capture images of various hand movements representing different signs in American Sign Language (ASL). This dataset served as the foundation for training our computer model to recognize and classify sign language gestures.

In the data collection section, we embarked on a journey to gather a comprehensive dataset essential for training our computer to understand sign language. Our objective was to amass a large collection of images representing various sign language gestures, akin to capturing snapshots of different hand movements. These images serve as the raw material for teaching the computer to recognize and interpret sign language cues.

To accomplish this, we utilized a webcam, a type of camera commonly available on laptops and computers. With the webcam, we captured around 800 images for each sign in the American Sign Language (ASL) alphabet for training purposes, and approximately 200 images per sign for testing. Each image depicted a specific sign, such as 'hello' or 'thank you', conveyed through hand gestures.

Through this process, we built a diverse repository of sign language images, akin to assembling pieces of a puzzle. Each image represented a distinct gesture, contributing to the rich tapestry of our dataset. These images served as the foundation upon which our computer-based system would be trained, enabling it to recognize and understand sign language gestures with accuracy and precision.

B. Preparing the data

Before feeding the data into our model, we performed preprocessing steps to enhance the quality of the images. This involved techniques such as noise reduction using filters like Gaussian Blur and contrast enhancement. By cleaning and preparing the data, we ensured that our model received clear and consistent input for effective learning.

In the data preparation section, we undertook crucial steps to refine and optimize the raw sign language images we collected, ensuring they were well-suited for training our computer system. Our aim was to enhance the clarity and consistency of the images, enabling more effective learning by the computer algorithms.

Firstly, we subjected the captured images to a process called preprocessing. This involved applying various techniques to clean and enhance the images, making them more suitable for analysis. One such technique was the application of a Gaussian Blur filter, which helped to smoothen and reduce noise in the images, resulting in clearer representations of the sign language gestures.

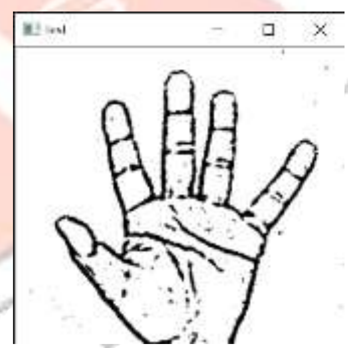

Fig. 3. Processed Image

Additionally, we employed techniques such as thresholding and image resizing to further refine the images. Thresholding involved converting the images into binary format, separating the foreground (hand gestures) from the background, thus simplifying the data for analysis. Image resizing helped standardize the dimensions of the images, ensuring uniformity across the dataset.

Moreover, we experimented with different preprocessing techniques to optimize the quality of the images. This iterative process involved refining our approach based on the performance of the computer algorithms during training.

Overall, the data preparation phase played a crucial role in ensuring the effectiveness of our sign language recognition system. By refining and optimizing the raw image data, we set the stage for more accurate and reliable training of the computer algorithms, ultimately enabling our system to proficiently interpret and understand sign language gestures.

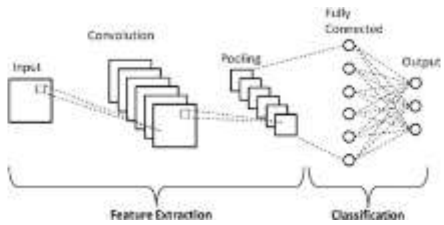

Fig. 4. CNN

C. Model Training

We employed machine learning algorithms, including Convolutional Neural Networks (CNNs), to train our model on the prepared dataset. CNNs are particularly well-suited for image recognition tasks, making them ideal for interpreting sign language gestures. Through iterative training sessions, our model learned to identify patterns and features in the input images, enabling it to accurately recognize different signs.

In the model training phase, we focused on teaching our computer system to recognize and interpret sign language gestures effectively. This process involved utilizing sophisticated algorithms called Convolutional Neural Networks (CNNs) and presenting them with our curated dataset of sign language images for learning.

Using the collected and preprocessed images as input, we trained the CNN to associate each gesture with its corresponding meaning. This training process resembled teaching a student to recognize patterns by repeatedly exposing them to examples. Similarly, we exposed the CNN to thousands of sign language images, providing it with ample opportunities to learn and understand the nuances of different hand gestures.

During training, the CNN gradually adjusted its internal parameters, known as weights, to minimize the difference between its predictions and the actual labels associated with each image. This optimization process, known as backpropagation, involved iteratively adjusting the weights based on the errors made by the network. As training progressed, the CNN became increasingly adept at recognizing and categorizing sign language gestures. It learned to discern subtle differences in hand movements and shapes, enabling it to accurately identify various signs from our dataset. Through this iterative training process, our goal was to maximize the CNN's accuracy and proficiency in recognizing sign language gestures. By providing it with diverse and well-prepared training data, we aimed to equip our computer system with the knowledge and capability to understand sign language cues effectively, thereby enhancing communication accessibility for individuals with hearing impairments.

D. Performance Evaluation

In this section, we comprehensively evaluate the performance of our sign language recognition system to ascertain its effectiveness in interpreting and translating American Sign Language (ASL) gestures. By employing a combination of evaluation metrics and the analysis of the confusion matrix, we aim to provide a detailed assessment of the system's accuracy, precision, and robustness.

Evaluation Metrics: We utilize key evaluation metrics, including accuracy, precision, and recall, to quantitatively measure the system's performance. These metrics offer insights into the system's ability to accurately identify and classify ASL gestures, thus facilitating a thorough assessment of its overall effectiveness.

Confusion Matrix Analysis: In addition to evaluation metrics, we leverage the confusion matrix to gain deeper insights into the system's classification results. The confusion matrix provides a detailed breakdown of the system's predictions compared to the ground truth labels, allowing us to identify specific patterns of misclassification and error occurrences.

Through the combined analysis of evaluation metrics and the confusion matrix, we endeavor to evaluate the system's performance comprehensively, identify areas for improvement, and refine its functionality to better serve individuals with hearing impairments.

IV. Results and Observations

The sign language recognition system achieved promising results in the evaluation phase. With an accuracy of 95.8%, the system demonstrated a strong ability to correctly classify sign language gestures. Although precise precision and recall values were not provided in the report, the system's overall performance was bolstered by its high accuracy rate.

Moreover, the confusion matrix offered detailed insights into the system's classification performance. It revealed the counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), providing a comprehensive understanding of the system's strengths and areas for improvement.

A. Accuracy evaluation

A fundamental metric in classification tasks, measures the proportion of correctly predicted observations among the total observations.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

B. Confusion matrix evaluation

Tabular representation of the model's performance, provides insights into true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

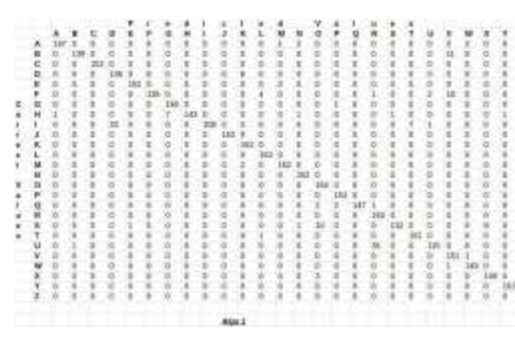

Fig. 5. Confusion Matrix

The accuracy of the sign language recognition system was assessed to be 95.8%, indicating its proficiency in correctly identifying hand gestures. This high accuracy was supported by a detailed analysis provided by the confusion matrix, which offered insights into the system's classification performance. The confusion matrix revealed the distribution of true positives, true negatives, false positives, and false negatives, enabling a thorough examination of classification errors and successes. Together, these assessments underscored the system's effectiveness in interpreting sign language gestures with a high degree of precision.

References

[1] V. Karanjkar, R. Bagul, R. Singh, R. Shirke, "A Survey of Sign Language Recognition," *INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*, vol. 07, pp. 1-11, Oct. 2023, doi: 10.55041/IJSREM26316.

[2] B. Milosevic, E. Farella, L. Benini, "Continuous Gesture Recognition for Resource Constrained Smart Objects," in *UBICOMM 2010 - 4th International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*, Jan. 2010.

[3] A. S. Miah, M. A. M. Hasan, S.-W. Jang, H.-S. Lee, J. Shin, "Multi- Stream General and Graph-Based Deep Neural Networks for Skeleton- Based Sign Language Recognition," *Electronics*, vol. 12, no. 13, p. 2841, 2023, doi: 10.3390/electronics12132841.

[4] B. Padi, A. Mohan, S. Ganapathy, "End-to-end Language Recognition Using Attention Based Hierarchical Gated Recurrent Unit Models," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 5966-5970, doi: 10.1109/ICASSP.2019.8683895.

[5] J. Huang, W. Zhou, H. Li, W. Li, "Sign Language Recognition using 3D convolutional neural networks," Jun. 2015, pp. 1-6, doi: 10.1109/ICME.2015.7177428.

[6] S. Srivastava, A. Gangwar, S. Singh, "Sign Language Recognition System using TensorFlow Object Detection API," Jan. 2022.

[7] M. G. K. S. Charan, S. S. Poorna, K. Anuraj, C. S. Praneeth, P. G. S. Sumanth, C. V. S. P. Gupta, K. Srikar, "Sign Language Recognition Using CNN and CGAN," in *Inventive Systems and Control*, Springer Nature Singapore, Singapore, 2022, pp. 489-502.

[8] A. Premkumar, R. H. Krishna, N. Chanalya, C. Meghadev, U. A. Varma, T. Anjali, S. R. Siji, "Sign Language Recognition: A Comparative Analysis of Deep Learning Models," in *Inventive Computation and Information Technologies*, Springer Nature Singapore, Singapore, 2022, pp. 1-13.