STM JOURNALS

Journal of
Electronic Design Technology

https://journals.stmjournals.com/joedt

ISSN: 2229-6980 (Online)
ISSN: 2321-4228 (Print)
Volume 15, Issue 2, 2024
May–August
DOI (Journal): 10.37591/JoEDT

Research JoEDT

Real-Time Gesture Recognition with Convolutional Neural Networks

Anjali Singh[1], Anjali Singh[1], Badal Singh[1], Deepa Shukla[1], Kirti Pandey*[1], A.K. Dohare[2], M. S. Mandloi[2]

Abstract

Sign language detection plays a pivotal role in bridging communication barriers for the deaf and hard of hearing community. An extensive investigation on the use of convolutional neural networks (CNNs) for sign language recognition is presented in this article. Leveraging the power of deep learning, our research aims to develop an accurate and efficient system capable of recognizing and classifying sign language gestures in real-time. The report begins with an overview of the significance of sign language detection in fostering inclusivity and accessibility. The approach is then covered in detail, including the architecture of the CNN model that was used, methods for gathering datasets and preparing them, and the training procedure. Results obtained from experimentation are presented, showcasing the effectiveness of the CNN model in accurately detecting a diverse range of sign language gestures. In conclusion, this report highlights the promising capabilities of CNN-based sign language detection systems, emphasizing their potential to enhance accessibility and communication for the deaf and hard of hearing community.

Keywords: American sign language (ASL), Sign language, hand gesture, convolutional neural network (CNN), machine learning, deep learning, OpenCv, Keras

INTRODUCTION

Sign languages are the foundation of local deaf cultures and have evolved into practical means of communication wherever there are deaf communities. While most users of signing are deaf or hard of hearing, hearing people also use signing, including those who cannot speak for physical reasons, those who struggle with oral language because of a disability or condition, and people who have deaf family members, including children of deaf adults. The practice of exchanging ideas and messages through words, gestures, body language, and visual aids is known as communication. People who are Deaf and Mute (Dumb) (D&M) utilize their hands to communicate with others by making various gestures. We understand the problems these people face and hence we tried to contribute to making their lives better by developing a system based on machine learning that will recognize sign language and convert it into text. This will help both hard of hearing and normal people to communicate with ease. Nonverbal cues are sent by gestures, which can only be perceived by eyes. It is necessary for two people to know and comprehend the same language to create communication. To converse, nevertheless, those who are deaf or hard of hearing require sign language. The relationship between spoken and sign languages is intricate and varies depending on the countries in which they are spoken. Standardized technologies are in place in several nations to interpret sign gestures.

***Author for Correspondence**
Kirti Pandey
E-mail: kirtipandey160501@gmail.com

[1]Student of Department of Electronics & Communication Engineering, Rewa Engineering College, Rewa, India
[2]Professor of Department of Electronics & Communication Engineering, Rewa Engineering College, Rewa, India

Received Date: July 16, 2024
Accepted Date: July 24, 2024
Published Date: August 04, 2024

Citation: Anjali Singh, Anjali Singh, Badal Singh, Deepa Shukla, Kirti Pandey, A.K. Dohare, M. S. Mandloi. Real-Time Gesture Recognition with Convolutional Neural Networks. Journal of Electronic Design Technology. 2024; 15(2): 12–18p.

There are various sets of sign languages like American sign language, Japanese sign language, Indian sign language, Chinese sign language, etc etera. Sign language is a nonverbal visual communication system that replaces spoken words with gestures, facial expressions, and body language [11].

It is used by deaf and hard-of- hearing individuals as well as those with speech or communication disorders to convey messages, ideas, and emotions. Sign languages are complete, natural languages with their own grammar, syntax, and vocabulary, and they are recognized as legitimate languages. We are here using American Sign Language as it is globally recognized as a standard sign language. ASL is taught in most of the educational institutes for deaf and mute people. One of the most widely used sign languages is American Sign Language (ASL). D&M people can only communicate through sign language because their only communication problem is connected to their inability to use spoken languages.

Hence, we decided to design a sign language recognition system which recognizes English alphabets and numbers. The gestures we trained are as given in the image below-Figure 1.

PROBLEM STATEMENT

Although there is many worlds' population who cannot speak and hear and hence depends completely on sign language to communicate with each other. But the problem arises while they try to communicate with people who don't know sign language properly. They face problems while conveying their messages and thoughts. To solve this problem, we are proposing a sign language recognition system, which will detect the hand gestures provided by verbally impaired person and will give the textual meaning of the gesture according to English sign language. The main goal of this project is to create a model that can identify hand motions based on fingerspelling and combine each gesture to generate a whole word.

With the help of embedded camera, the model will capture the real time images of the Gestures shown by the person and will give the text output of the meaning of the gesture shown. Hence it will make the conversations between a speaking impaired person and a normal person a lot easier and convenient.

LITERATURE REVIEW

In recent years, sign language recognition has drawn a lot of interest from researchers. Numerous scholars have worked in this field to reduce the communication gap that exists between the deaf and the non-deaf. Numerous approaches have been attempted to maximize the system's performance.

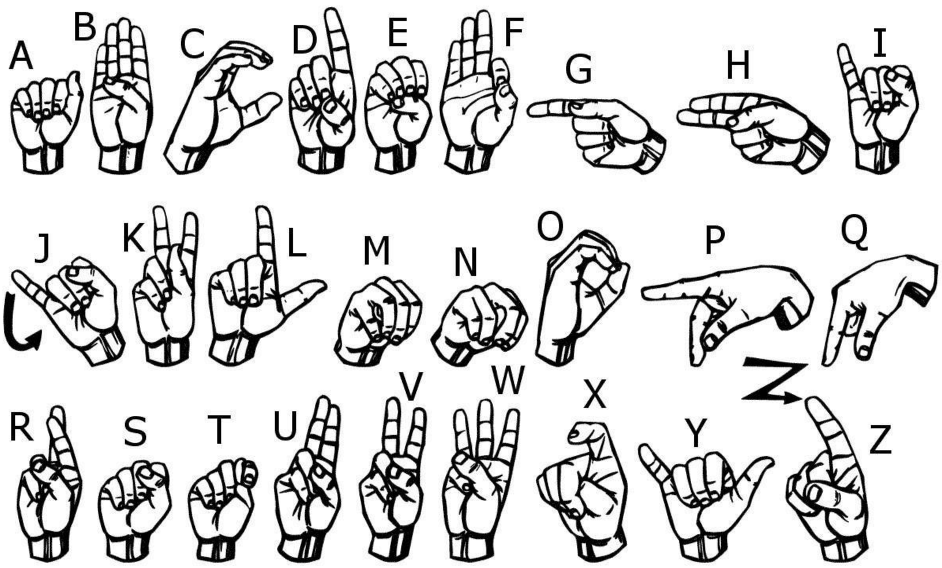

Figure 1. ASL Hand Gestures.

There are two methods for detecting sign language: the glove-based method, which is more accurate, and the vision-based method, which is more sensitive to setup and movement restrictions. The vision-based method consists of two steps: the first is feature extraction, which extracts static frames from videos using image processing, and the second is recognition, which learns by finding patterns in training data and tracking.

In the paper by Ms. Greeshma Pal et.al [1] described the uses of CNN, KNN and SVM for hand sign recognition. They performed a comparison-based study on these methods and according to that they proposed a statement that KNN has an accuracy rate of 93.83% and, SVM has an accuracy rate of 88.89% which is lower than KNN meanwhile CNN was proven as the most accurate method for hand gesture recognition as it has the highest accuracy of 9849%. That is why we are using CNN in our project. Another paper by Lionel Pigou et.al [2] assures about the accuracy of CNN in Sign Language Recognition. Image Processing and Deep Learning can also be infused to get desired results. A real-time model for ISL gesture detection utilizing image processing and deep learning is proposed in a study by Neel Kamal Bhagat et al. [3]. They use the incoming image data from Kinect and then with the help of depth perception and computer vision techniques they perform background subtraction and one to one mapping between depth and RGB pixels. To identify the clear hand gesture. This method is also highly accurate. Another method used is Principal Component Analysis (PCA) and Artificial Neural Network (ANN) algorithms used in MATLAB.

In paper proposed by Kusumika Krori Dutta [4], they showed that single and double-handed Indian Sign Language (ISL) can be acquired for all English alphabets using KNN and Back Propagation techniques where PCA is used3 fir dimension reductions. It had an accuracy rate above 96%. Rachna Patil et.al [5] also used CNN for Indian Sign Language Recognition and achieved accuracy rate of 95%. Ying Xie and Kshitij Bantupalli et.al [6] took to CNN and used American Sign Language (ASL) dataset. They also used it create vision based ASL recognition model. Japanese fingerspelling recognition system by N. Mukai et.al [7] is based on classification tree and machine learning. It makes use of Japanese fingerspelling hand gestures. It uses Support Vector Machine (SVM) as a machine learning method. Also, the Chinese fingerspelling recognition system is used. It uses Microsoft Kinect. They utilize moving trajectories obtained by Kinect. T. Liu on his paper et.al [8] described this technology. There is another method used for recognizing Chinese sign language. Its framework is based on Hidden Markov Model (HMM), and they extract motion trajectory feature from video sequence. It's explained in detail in paper cited by J. Zhang et al [9]. There had been numerous attempts to detect hand gestures or signs and combine it with different techniques to get accurate results. Even different countries have different set of gestures for their native users.

METHODOLOGY

Steps involved for creating this project is as follows:

Data Collection

A comprehensive dataset of sign language gestures is compiled. This includes images or video frames of different signs performed by various individuals to account for variations in skin tone, hand size, and lighting conditions.

Preprocessing

Images are resized and normalized to ensure uniformity. Data augmentation methods such as flipping, zooming, and rotating are used to improve the resilience of the model.

CNN Architecture

A deep CNN is designed, typically consisting of multiple convolutional layers followed by pooling layers to reduce dimensionality. The completely linked layers at the end of the network produce probabilities for every sign class.

Training

Using an optimizer like Adam and a loss function like categorical cross-entropy, the model is trained on the gathered dataset. To stop overfitting, regularization strategies like dropout are employed.

Evaluation

Metrics including accuracy, precision, recall, and F1-score are used to assess the model's performance on a validation set. Fine-tuning is done based on the evaluation results to improve accuracy.

Deployment

The trained model is deployed in a real-time application where it captures live video, processes frames, and detects gestures.

SYSTEM FLOW

Image Capture

Prepare the webcam of our system to operate through program.

Hand Detection & Tracking

In this step, the hand detection takes place in video feed from our webcam.

Data Acquisition

In this step, we click images from webcam and create a database which will later be used to train the model for detecting hand sign gestures. One of the examples is shown in Figure 2. Image of Letter “A” stored in Database.

Convert Into Grey Scale

This step reduces the size of image without losing any information by using Grey Scale filter (Figure 3). It only stores the outline of the image and only those parts of images which are enough to distinguish one gesture from other, so that no confusion can appear while recognising user’s hand gestures. This property helps in reducing the size of the data stored in the device.

Segmentation

In this step, the entire image is divided into small segments for further processing of the image. It also consumes less storage space and increases speed of operation.

Feature Extraction

This step extracts features from our image. We are using the Convolutional Neural Network (CNN) to extract different features from our images.

Training Database/Model

In this step we train our model for sign language recognition.

Testing Code

In this step, we test our codes and model performance to provide some result.

Extract Output

In this step, we get the required output. Our given input is in image form and given output is in text form.

Customization

In this step the output text will get modified, if it matches with the word or task we were looking for, then we can get the actual output of our need.

The system flow diagram of the proposed system in shown in Figure 4.

ISSN: 2229-6980 (Online), ISSN: 2321-4228 (Print)

Figure 2. Image of Letter 'A' Stored in Database.

Figure 3. Gray Scale Image of Letter 'A'.

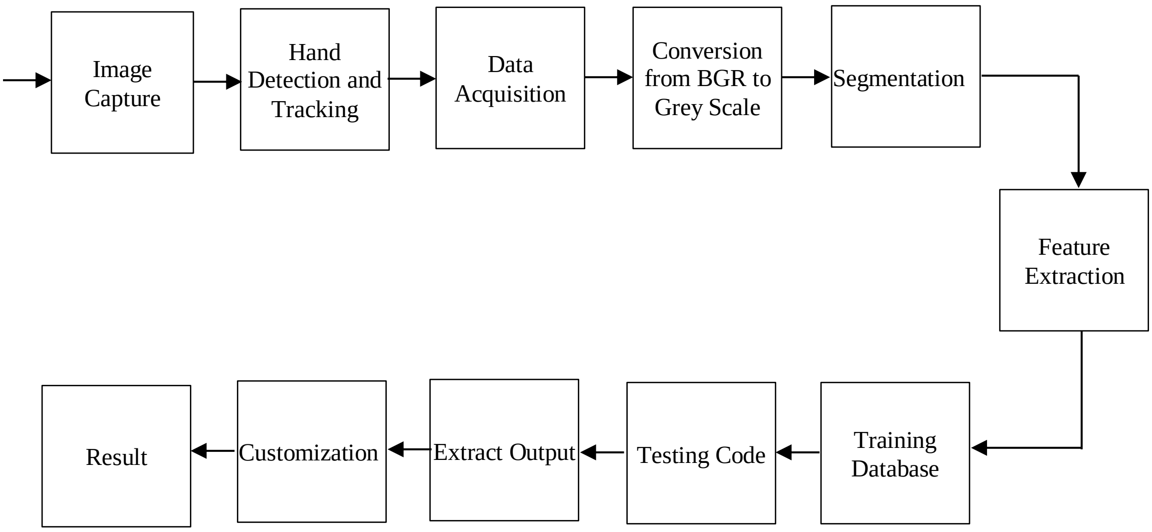

Figure 4. System Flow Diagram of the proposed system.

RESULT

The CNN model shows excellent recognition accuracy for a large variety of sign language movements. The use of data augmentation and regularization techniques significantly enhances the model's generalization capabilities. The system provides real-time translation of sign language into text, showing potential for practical applications.

CONCLUSION

The application of CNNs in sign language detection presents a promising solution for improving communication accessibility for the deaf community. Future work includes expanding the dataset, incorporating more complex gestures, and improving the system's real-time processing capabilities. However, we also acknowledge several areas for future improvement and exploration. Further optimization of network architectures, fine-tuning hyperparameters, and expanding the dataset size and diversity could lead to even higher accuracy and generalization capabilities. Additionally, exploring multimodal approaches, incorporating in-depth information, and addressing real-time performance challenges are avenues worth pursuing.

Future Scopes

Neural Networks (CNNs) are quite promising and diverse. Here are several potential avenues:

1. *Dataset Expansion:* Including more signs, facial expressions, and contextual variations to enhance the model's versatility.
2. *Integration with Other Technologies:* Combining with natural language processing (NLP) for better interpretation of context and meaning.
3. *Improving Real-Time Performance:* Optimizing the model for faster inference times to ensure smooth real-time communication.
4. *Improved Accuracy and Efficiency:* Researchers will likely continue to focus on enhancing the accuracy and efficiency of CNN models for sign language detection. This could involve developing more sophisticated network architectures, optimizing training algorithms, and leveraging larger and more diverse datasets.
5. *Cross-Lingual Sign Language Recognition:* As sign languages vary across different regions and cultures, there's a need for sign language detection systems that can recognize and interpret signs from multiple languages. CNNs can be trained on multilingual datasets to achieve cross-lingual recognition.
6. *Adaptation To New Environments:* Sign language detection systems should be able to adapt to different environments and lighting conditions to maintain robust performance. Future research may focus on developing CNN models that are more resilient to variations in input data.
7. *Privacy and Ethical Considerations:* Sign language detection systems should be able to adapt to different environments and lighting conditions to maintain robust performance. Future research may focus on developing CNN models that are more resilient to variations in input data.
8. *User-Friendly Interfaces:* Designing intuitive and user-friendly interfaces for sign language detection systems is crucial for ensuring accessibility and usability. Future research may explore novel interaction paradigms, such as gesture-based interfaces or augmented reality applications.

Overall, the future of sign language detection systems using CNNs is bright, with ongoing advancements aimed at improving accuracy, speed, versatility, and accessibility.

REFERENCES

1. G. Pala, J. B. Jethwani, S. S. Kumbhar and S. D. Patil, “Machine Learning-based Hand Sign Recognition,” 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 2021, pp.356-363, doi:10.1109/ICAIS50930.2021.9396030.
2. Lionel Pigou, Sander Dieleman, Pieter- Jan Kindermans and Benjamin Schrauwen, “Sign Language Recognition Using Convolutional Neural Networks,” 2015 European Conference on Computer Vision.
3. N. K. Bhagat, Y. Vishnusai and G. N. Rathna, “Indian Sign Language Gesture Recognition using Image Processing and Deep Learning,” 2019 Digital Image Computing: Techniques and Applications (DICTA), Perth, WA, Australia, 2019, pp. 1- 8, doi: 10.1109/DICTA47822.2019.8945850
4. Dutta KK, Bellary SA. Machine learning techniques for Indian sign language recognition. In2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC) 2017 Sep 8 (pp. 333-336). IEEE.
5. Patil R, Patil V, Bahuguna A, Datkhile G. Indian sign language recognition using convolutional neural network. InITM web of conferences 2021 (Vol. 40, p. 03004). EDP Sciences.

6. K. Bantupalli and Y. Xie, “American Sign Language Recognition using Deep Learning and Computer Vision,” 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 4896- 4899, doi: 10.1109/BigData.2018.8622141.
7. N. Mukai, N. Harada and Y. Chang, “Japanese Fingerspelling Recognition Based on Classification Tree and Machine Learning,” 2017 Nicograph International (NicoInt), Kyoto, Japan, 2017, pp. 19-24, doi: 10.1109/NICOInt.2017.9.
8. T. Liu, W. Zhou and H. Li, “Sign language recognition with long short-term memory,” 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 2016, pp. 2871-2875, doi: 10.1109/ICIP.2016.7532884.
9. J. Zhang, W. Zhou, C. Xie, J. Pu and H. Li, “Chinese sign language recognition with adaptive HMM,” 2016 IEEE International Conference on Multimedia and Expo (ICME), Seattle, WA, USA, 2016, pp. 1-6, doi: 10.1109/ICME.2016.7552950.
10. P. C. Badhe and V. Kulkarni, "Indian sign language translator using gesture recognition algorithm," 2015 IEEE International Conference on Computer Graphics, Vision and Information Security (CGVIS), Bhubaneswar, India, 2015, pp. 195-200, doi: 10.1109/CGVIS.2015.7449921
11. H. Adhikari, M. S. Bin Jahangir, I. Jahan, M. S. Mia and M. R. Hassan, "A Sign Language Recognition System for Helping Disabled People," 2023 5th International Conference on Sustainable Technologies for Industry 5.0 (STI), Dhaka, Bangladesh, 2023, pp.1-6,doi: 10.1109/STI59863.2023.10465011.

